**Three Essays in the Economics of Migration**

by

Alexander D. C. Persaud

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2017

Doctoral Committee:

Professor Paul W. Rhode, Chair
Professor Achyuta R. Adhvaryu
Professor C. Hoyt Bleakley
Professor Charles C. Brown

Alexander D. C. Persaud

persauda@umich.edu

ORCID iD: 0000-0003-4712-197X

## DEDICATION

I dedicate this dissertation to my loving family in all parts of the world and to all the immigrants, past and present, who have made the decision to leave their homes.

ACKNOWLEDGMENTS

First, I thank my committee for their assistance in completing my research: Paul Rhode (chair), Achyuta Adhvaryu, Hoyt Bleakley, and Charlie Brown. Second, I thank Raj Arunachalam for his support and mentorship in my first three years of graduate school. Third, I thank the Economics Department staff for their support and assistance during my five years.

For my research on historical Indian migration, I thank many people for comments and research advice: Martha Bailey, John Bound, Latika Chaudhary, Christian Dippel, Prachi Jain, Ajin John, Gaurav Khanna, David Lam, Suresh Naidu, Samuel Norris, Kevin O'Rourke, Ajay Shenoy, Jeff Smith, Mel Stephens, Andrew Usher, Gonzalo Vasquez-Baré, Tejaswi Velayudhan, Dean Yang, and many seminar participants at the University of Michigan, the NBER DAE summer institute, Princeton University (EconCon), the Economic History Association, Oxford University, NEUDC, Carleton College, and the Economic History Society. Anuraag Aekka, Shreyas Bindal, Michael Henry, Akhila Kadiyala, Roman Klimke, Heather Smallwood, and Varsha Swamy provided excellent research assistance. Staff at the National Library of Australia, the National Archives of Guyana, the National Archives of Jamaica, the National Archives of Trinidad and Tobago, and the University of KwaZulu-Natal assisted in record location. I thank the Michigan Institute for Teaching and hResearch in Economics (MITRE) and the University of Michigan Rackham Graduate School for research trip funding.

For my research on contemporary German migration and labor markets, I thank Daniela Hochfellner, the staff at the Research Data Centre (FDZ) of the Federal Employment Agency at the Institute for Employment Research (Das Forschungsdatenzentrum der Bundesagentur für

Arbeit im Institut für Arbeitsmarkt- und Berufsforschung), and the staff at the University of Michigan's Institute for Social Research for various assistance in my projects. I also thank seminar participants at the University of Michigan, the First International User Workshop of the FDZ, the Eastern Economic Association, and the Western Economic Association.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation covers three important topics in international migration: why people migrate, why some people return home after initially migrating, and what happens in the labor market to those who stay. I utilize new, unique microdata on roughly 250,000 Indian indentured servants sent around the world in the nineteenth and early twentieth centuries under forward contracts for the first two chapters. In the third chapter, I rely on administrative data from West Germany/Germany from the 1970s to early 2000s.

Uncertainty about economic conditions, not merely average wage differentials between markets, affects migration. Ex ante forward, guaranteed migration contracts—used by millions of workers—can reduce uncertainty for migrants. In the first chapter, I combine aspects of the two and ask, how does origin-market uncertainty affect out-migration under forward contracts?. Indian indentureship offers a clean identification strategy to isolate the importance of sending-market uncertainty. The migration decision is consistent with migrating to escape price volatility, my main measure of uncertainty. I find compositional differences by social network: different castes respond differently to economic conditions. The effects of prices and wages on landowning castes are the reverse of those on non-landowning castes. Finally, volatility exerts a persistent negative effect by lowering return migration.

The second chapter further develops the research on return migration. I ask, how much are upper-caste individuals willing to pay for upper-caste status? Traditional discrimination models view discrimination from the vantage point of a group that receives worse treatment due to its non-economic characteristics vis-a-vis a reference group. Discriminated groups pay a cost

in the labor market for these characteristics. However, the corollary may be true: a privileged group may pay a fee in order to receive better, rather than equal, treatment. In the case of caste, I hypothesize that high-caste Indians abroad are willing to pay more to return to India and reap the labor and non-labor benefits of high-caste status. To test this, I use return migration of Indian indentured servants in Fiji. This context removes confounding labor-market factors and cleanly identifies the gross value of upper castes. I show that the lower bound of the value of the highest castes in north India roughly 2.5 years' gross wages. The effects are entirely driven by men, as women's caste status appears delinked from return migration. My results show some of the first evidence quantifying a caste's value and speak to caste's persistence today.

In the third chapter I switch to contemporary West Germany/German to study labor market differences after migration for natives and immigrants. Using a representative panel data set of workers, I find that differentials do exist between German workers and foreign workers. Human capital accounts for many of the differences, with both differential levels and returns for each group. First, natives are more skilled than foreigners. Second, at low levels of education and skill, foreigners earn roughly 4% more than natives; however, this reverses to foreigners earning 3% less at high levels of education and skill. German and foreign workers also exhibit different wage-growth trends over their work cycles, with foreign workers earning on a flatter trajectory. Return migration exacerbates these trends. Using an alternative administrative data set and comparing, I conclude that return migration increases the wage differential over time due to positive selection of return migrants.

**CHAPTER 1**

**Risk mitigation and selection under forward contracts: 19th-century Indian indentureship**

# 1 Introduction

Millions of workers migrate around the world under forward contracts as guest workers or another form of bonded laborers. For example, millions of *Gastarbeiter* (guest workers) entered West Germany from the 1950s - 1973.[1] Many developed countries maintain guest worker programs today such as the US temporary worker programs under the H2-A and H2-B visas, and there are calls to expand these visas.[2] The *kafala* system in the Middle East ties many foreign workers to an employer for a specified period of time. Previous scholarship has studied remittances, return migration decisions, and the effects of recipient-market conditions such as minimum wages or employer transitions (e.g., Naidu et al. 2015, Yang 2006, McKenzie et al. 2014). However, *why* people select these contracts remains understudied.

Another field of scholarship emphasizes the important roles of volatility in economic outcomes as well as the relationship between volatility and migration. Jayachandran (2007) shows that volatility due to productivity shocks and poor insurance markets dampen wages and raise the returns from migrating. Her margin of adjustment is local labor force supply, and she leaves migration as an open question. However, uncertainty in the recipient market and an inability to insure against this uncertainty may prevent migration even when there are high returns (Bryan et al. 2014). Uncertainty may break a competitive wage market and lead to share-cropping (Stiglitz 1974).

[1] Honekopp 1997.

[2] See Porter, Eduardo. "If Immigration Can't Be Stopped, Maybe It Can Be Managed." *New York Times* 26 October 2016, page 81.

Volatility is an important concept in other fields. In international trade, exchange rate volatility is studied because of its potential impacts on trade flows due to uncertainty. Policy uncertainty can harm the overall economy by causing drops in investment, output, and employment (Bloom 2009, Baker et al. 2015). I combine the labor-market structure of guest workers and the conceptual issue of volatility and ask, how does origin-market uncertainty affect out-migration under forward contracts?

To my knowledge, this paper is the first to study the general topic of origin-market uncertainty on migration, as well as the use of forward contracts in particular. I turn to a setting with guaranteed forward contracts abroad and uncertainty in the home market in order to identify cleanly the role of uncertainty: Indian indentureship in the late nineteenth/early twentieth centuries. Over one million Indians became indentured servants to colonies as far away as Jamaica, South Africa, and Fiji over several decades, which provides variation in time, space, and contract. Volatility in prices and wages created uncertainty about economic outcomes in India. On the recipient side, indentureship contracts lasted for several years, specified a wage in advance for the duration of the contract, and covered transportation costs. This last provision reduced or eliminated the barriers to migration for workers living near subsistence.

I hypothesize that Indians chose to use indentureship as a risk-mitigation mechanism to smooth consumption over time in response to local (i.e., Indian) uncertainty. Furthermore, I hypothesize that variation in Indian uncertainty, measured here as price volatility, induces much of the selection and variation in the indentureship population. Figure 1 graphs price volatility in India against total emigration under indentureship contracts along with polynomial fits for both series. The two track each other well and provide suggestive evidence for my hypothesis.

My approach differs from migration work in developed countries (e.g., Kennan and Walker 2011) and also developing ones. Building off the canonical work from Harris and Todaro (1970) and Borjas' use of the Roy Model (1987), I develop a migration model that includes uncertainty, risk aversion, and intertemporal decision-making. The model highlights the long-run nature of indentureship and risk-mitigation, which previous models have omitted. Risk aversion provides a new parameter of interest, sending-market consumption volatility, that enables me to examine how both the first and second moments of consumption affect migration. The results are generalizable to two markets with volatility.

I collect unique microdata on roughly 250,000 Indian indentured servants to multiple colonies to test the model's predictions. I also collect aggregate statistics from Indian districts and several colonies, including data newly digitized for this project. I trace individuals to their home districts and map on prices, wages, and measures of volatility to create a district-level panel dataset spanning several decades. I model out-migration by district. Next, I turn to the individuals' characteristics and test for effects of prices, wages, and volatility on the composition of the indentured servants. Finally, I trace out how initial volatility affected return migration in a context where settlement and return migration were both facilitated.

My findings are broadly consistent with the use of indentureship as a risk-mitigation device against volatility. First, volatility significantly raises the probability of a district sending a migrant. Second, volatility operates on the intensive margin, too, and raises the number of indentured servants from a given district. Third, there are heterogeneous price and volatility effects by social network, for which I use caste as a proxy. Upper castes and other castes respond differently to prices, but lower castes migrate more in the face of volatility.

My research speaks directly to migration today in the Middle East under the *kafala* system. The long contract lengths, fixed periods of time, negotiated contracts, and international aspect of indentureship is the historical analogue to *kafala*. Given the paucity of data and empirical research, with the notable exception of Naidu et al. (2015), my research provides answers to topics of contemporary relevance to policymakers in countries like India and Nepal, from which many workers go to the Middle East. My results show that policymakers must take more than wage levels into account but also volatility of income, something partly addressed in the last decade in India with NREGA.

The limited past research on Indian indentured servants has studied the characteristics of migrants rather than their causes of becoming indentured servants. Brennan et al. (1994, 1998, 2003) use data from Fiji and other colonies to quantify changing heights over time. However, they do not examine the why or who questions and focus on anthropometrics. Furthermore, they fail to account for sample selection while drawing larger conclusions about Indian health/wellbeing in the nineteenth century. Galenson (1984) focuses on the provision of transportation to the West Indies but does not examine how sending-market characteristics affect take-up into indentureship or varying transportation costs to other destinations closer to India.

Prior research on another form of indentureship, that to colonial English America, has focused on contractual terms due to the vast institutional differences that allowed for person-specific contracts. Galenson (1981) studies indentured servants in colonial English America and finds that indentureship contracts varied with human capital and other characteristics, which shows a market response to the potential outside options of immigrants. Abramitzky and Braggion (2006) study inter-colony choice and human capital in colonial English America and find evidence of positive selection amongst migrants to the American mainland. However, the lessons from colonial America differ from the Indian case due to different contract types and post-indentureship economic options. Early American indentureship allowed for more market adjustment and variation in contract due to less regulation. Indian indentureship, in contrast, provided a fixed, known contract and contracting party. As noted above, Indian indentureship bears more similarities to modern-day forms of bonded labor than to this historical form.

Research on a similar period of time in the British Empire studies different questions apart from indentureship and volatility. Naidu and Yuchtman (2013) examine master-servant contracts, a form of bonded labor, in industrial Britain during the late nineteenth century. They find that such contracts kept wages low and less volatile. However, there is no migration aspect to this, and master-servant contracts were vestiges of older apprenticeship systems and not a new system of long-term contracts. Dippel et al. (2015) touch on indentureship in their study the effect of commodity price shocks and post-emancipation labor dynamics in British West Indian colonies. However, they focus on intra-West Indian political economy and not on the Indian side of indentureship.

Historians have studied Indian indentureship but largely from a qualitative perspective that emphasizes the political economy of the British Empire and the need to provide labor for plantation economies in post-slave societies. Examples of research on Indian migration include Adamson (1973), Tinker (1974), Rodney (1981), Mangru (1987), and Seecharan (1997). Lal (2004) stands out as a notable exception. He quantifies the numbers of Fijian indentured immigrants and provides a broader overview of indentureship. However, he explicitly eschews statistical modelling and cannot make tighter conclusions beyond the general patterns observable in the data.

In addition to highlighting the role of volatility, my research contributes to three salient topics in development economics and economic history: social insurance, public policies for agricultural areas, and south-south migration. First, I offer further evidence on the role of social network, particularly caste, in providing social insurance. Rosenzweig and Stark (1989) find that migration through marriage is used as a form of co-insurance between extended families. Morten (2014) models the joint decision of co-insurance and temporary migration and finds that the two substitute for one another. My results move into labor markets and are consistent with Munshi and Rosenzweig (2016) in that different castes make different migration decisions. This may result from caste-specific insurance. I find that Indians of many castes were mobile in the late nineteenth/early twentieth centuries. This raises questions about the changing salience and meaning of caste in the modern era. Second, this research speaks to policies to stabilize consumption-good prices or to offer insurance in agricultural settings. Such policies may reduce demand for bonded-labor options. Third, I study global south-south migration, which, as Hattan and Williamson (2002) point out, "is not new. It is just ignored by economists." I offer new insight into this understudied phenomenon.

## 2 Background

As historians have shown, the political economy of the British Empire created the institution of Indian indentureship. The British Empire abolished the slave trade in the early 1800s, and emancipation followed with the Slavery Abolition Act of 1833.[3] Planters in some colonies abandoned sugarcane, a primary, labor-intensive crop under slavery. Although slaves were required to serve an additional period of apprenticeship at the behest of planters, former slaveowners in the West Indies and other colonies turned to other methods of guaranteeing and controlling the labor supply. In Barbados, for example, planters relied on former slave labor and restricted out-migration in order to maintain low wages (Roberts 1955). Other colonies turned to outside labor, and India became the prime supplier of labor. Sugarcane was the primary crop in colonies that imported indentured labor through the end of the nineteenth century.

The Indian and recipient colony governments set up indentureship contracts by law, which facilitates the collection of data and analysis of migration decisions. Government-sponsored indentured immigration supplanted free migration in order to satisfy labor demands while creating credible contracts. Potential Indian migrants to participating colonies were either required to pay the passage themselves or sign on to government-overseen indentureship contracts. Private con-

[3]Emancipation for most slaves was not immediate, with many forced to serve an additional apprenticeship until 1 August 1838.

tracts, which typified colonial English American indentureship, were illegal. Hundreds of licensed recruiters spread around India enrolled workers on a rolling basis throughout a year. Unlike similar migration to Assam's tea estates, very few individuals went as paid passengers.[4] The small number of free passengers bypasses problems with selection into alternative forms of migration. Recruitment for tea estates and overseas indentureship largely did not overlap, with tea focused on peripheries in the east and northeast. Only Bombay, Calcutta, and Madras were allowed to serve as entrepôts for emigration, and Bombay remained a minor port overall.

Contracts specified the length of time, usually five years; a fixed wage for the categories of men, women, and children; food, housing, and medical benefits; and the provision of additional benefits at the end of the contract. Figure 2 shows a contract (in Devanagari) with the wage underlined. Remarkably, the contracts were of long duration and there were only minor legislative changes over time in pay. Contracts did vary across colonies, though. Like master-servant contracts in the United Kingdom proper, breach of the indentureship contract was a criminal, not civil, offense.

Over one million Indians eventually became indentured servants in a variety of locations from 1838-1917.[5] Mauritius was the first colony to experiment with importing Indian labor, and the practice spread to other British colonies. Natal and Fiji stand out as later primary destinations. Several French colonies, including Cayenne (French Guiana), Guadeloupe, Martinique, and Réunion, adopted indentureship, although many migrants departed not from British India but rather from French India. Dutch Guiana (Suriname) and the Danish island of St Croix also petitioned British officials for indentured migrants and were permitted to do so, although the latter received only one ship and discontinued the practice within a decade. Officials in India required the Dutch to build a similar indentureship bureaucracy in Dutch Guiana as in British colonies with a protector in Paramaribo and sub-agents, and British officials in Trinidad oversaw Danish indentureship. Table 1 gives an overview of the years of indentureship and statistics.

Fiji, Natal, and Suriname form the primary focus of this paper due to the preservation and digitization of their records and the large proportion of indentured servants they received from 1873-1916, the years that overlap with Indian price and wage data. I add several years of Jamaican

[4]For example, the data that I collect for Fiji show 0.7% of all Indians went as paid passengers.

[5]In comparison, it is estimated that around six million Indians travelled under various forms of unregulated contracts to outlying areas of British India (Burma, Ceylon, and Malaya) from the mid-nineteenth century through 1938.

data, too, from some of the only records to survive. Fiji and Natal furthermore provide wide time and geographic variation: they drew from a rich mix of Indian districts via Madras and Calcutta for roughly 40 years. In Fiji and Natal, Indian labor was essential to starting the indigenous sugar industry. In Jamaica and Suriname, Indians replaced emancipated slaves in the sugarcane estates.

The growing home rule movement in India viewed indentureship negatively as an institution designed to exploit Indian labor for low wages. Indentureship was halted in 1917 due to political pressure in India, and the last indentured immigrants completed their contracts in 1922. Despite this, thousands of Indians continued to sign on as indentured laborers in the mid-1910s up until the government-imposed cessation.

# 3 Model

I modify earlier migration models to include risk aversion, forward contracts, and uncertainty with two periods. I use market 0 and market 1 as the home and recipient markets, respectively, to maintain consistency with Borjas (1987). I highlight the roles of three important actors: laborers in market 0, who may migrate to market 1; firms in market 1; and the government.

The use of uncertainty and loss aversion poses different implications than the Harris-Todaro Model. In that model, expected wages in their urban sector depend on the probability of unemployment, and the equilibrium condition equalizes the unemployment-adjusted wage with the rural wage. Here, though, the equalization of expected consumption bundles (wages) across sectors does not lead to equilibrium; rather, the equalization of expected utility does. Though this point seems subtle, the underlying mechanism is different: workers in my model face sending-market uncertainty and are risk averse, which in turn spurs migration.

## 3.1 Worker's problem

Laborers in market 0 live in $J$ districts operating under a traditional economy subject to fluctuations in real wages. I assume log utility for risk aversion.

Initially, in a closed economy without migration, worker $i$'s utility in time $t \in \{1, 2\}$ and district $j$ is given by

$$u_{ijt} = U(c_{ijt}, A_j)$$

where $U$ is a twice-differentiable, concave function, $c_{ijt}$ is consumption for person $i$ in district $j$ at time $t$, and $A_j$ represents a measure of local amenities distinct to village $j$. By these properties, $U_1 > 0, U_{11} < 0, U_2 > 0, U_{22} < 0$. The cross-partial $U_{12} \geq 0$. There is full support for the $A_j$ term in $\mathbb{R}$.

For exposition, the worker is assumed to supply labor inelastically, so leisure does not enter the utility function except as part of the general amenity term. The worker earns a real wage $w_{ijt} = w_{jt} > 0$ via subsistence production in each period. The distribution of wages is given by $\log(w_{jt}) = \mu_j + \epsilon_{jt}$, where $\epsilon_{jt} \sim \phi(0, \sigma^2_{w_j})$, where I impose no assumptions on $\phi$ except that it is not a degenerate function.

Workers live hand-to-mouth, and there is no storage of the consumption good or wages. There is free disposal of wages within a period. The consumption good price is normalized to 1 within each district. The maximization problem immediately simplifies to $c_{ijt} = w_{ijt} = w_{jt}$. Note that $\sigma^2_{w_j} = \sigma^2_{c_j}$.

Now suppose that the worker can choose at the start of period 1 to migrate to a foreign labor market under a credible, guaranteed forward contract. The real wage of this contract is $\tilde{w}$ and translates into $\tilde{c}$ consumption, defined ex ante, and the local amenities are given by $A_f$ where $0 < A_f \leq A_j$ for all villages $j$.[6] Mechanically, $d\tilde{c}/d\tilde{w} > 0$. Because the labor market is foreign, wages in the foreign market are independent of wages in any given village, i.e., $\forall j, \tilde{w} \perp w_j$. Assuming that the labor supply remains fixed across labor markets, the migration decision for $i$ comes from the inequality

$$\begin{aligned}
&(1+\beta)U(\tilde{c}_j, A_f) \lesseqgtr U(c_{j1}, A_j) + \beta E[U(c_{j2}, A_j)] \\
&\Leftrightarrow \underbrace{\beta^{-1}(U(\tilde{c}, A_f) - U(c_{j1}, A_j)) + U(\tilde{c}, A_f)}_{\equiv \pi} \lesseqgtr E[U(c_{j2}, A_j)]
\end{aligned} \tag{1}$$

where $\beta < 1$ is the discount factor.[7] The worker observes the wage level in the village economy, $\tilde{w}$ internalizes the monetary costs of migration[8], and there is no return migration (due to the forward-

[6] This assumption can be relaxed but serves to give a more conservative bound on the migration decision.

[7] This resembles Massey et al's (1993) basic formulation of returns except in discrete time with a guaranteed return over the worker's earnings horizon.

[8] This can be relaxed with a cost $C_j$ specific to each village $j$. In that case, $c_f$ is not constant and instead varies by the origin $j$, but the remaining algebra goes through with an additional comparative static that migration increases from $j$ as the cost $C_j$ decreases. Alternatively, variation in cost can be captured in the relative difference of amenities.

looking, multi-period nature of the foreign contract and the two-period lifespan). If $U$ is additively separable between its arguments, then the above problem simplifies to an expectation just over $c_{j2}$ on the right hand side.

By concavity, if $U(E[c_{j2}], A_j) \leq \pi$, then migration definitely occurs. Moreover, the worker will migrate even for some values where $U(E[c_{j2}], A_j) > \pi$ if there is any uncertainty.

Equation 1 leads to several immediate comparative statics:

- $d\pi/d\tilde{w} > 0, d\pi/dA_f > 0$: higher wages or amenities in the foreign market raise the value of $\pi$ and increase the probability of migrating

- $d\pi/dw_j < 0$: higher wages in the origin market raise the current value of staying and reduce the probability of migrating

- $dE[U(c_{j2}, A_j)]/d\sigma_{c_j} < 0$: higher variation in consumption, which captures uncertainty, lowers the expected utility of remaining in village $j$ and increases the probability of migrating

With $\sigma_{c_j} > 0$, it is possible for local wages to exceed foreign ones and yet have migration. Intuitively, in the local market, greater wage variation induces uncertainty about future wages, which makes the certainty value of the foreign contract more appealing.

I assume that agents form an expectation of $\sigma_{c_j}$ based on their past observations. This form of adaptive expectations means that higher fluctuations in consumption in a market $j$ in the past imply further fluctuations in the future.[9]

The upshot of the model seems to clash with Turnovsky et al. (1980) and earlier work on price volatility, in which price stabilization (or certainty) hurts consumers. However, in broad cases in which consumers spend large parts of their income on a good or risk aversion is high, price stability is beneficial. Both of these conditions hold here: the worker consumes all of her income on a consumption good, and the worker is risk averse. Thus, the model's predictions line up with earlier theoretical work on price variations.

This model distinguishes itself from Borjas' use of the Roy Model in three major ways. First, the forward wage incorporates migration costs from the decision. As noted above I incorporate differences in valuation between the two markets, which encompass psychological migration costs, into the amenity terms.

[9] I show that this is the case empirically for my context below.

Second, the ex-ante wage contract creates a unit mass distribution in the foreign labor market wage. I remove variation from one side of the migration decision. The skills, productivity shocks, and other factors in the home market that may lead to different wages across different $i$ are flattened. The correlation $\rho_{w_j,\tilde{w}}$ between the wage at home and abroad becomes immaterial (and zero) because the fixed $w_f$ makes the value of migrating, $\pi$, known ex ante.

Third, the migration decision depends on an expectation of future income. The worker compares the foreign offer against the local distribution of wages and, based on a belief about the subsequent year's wages, decides. This lack of ex ante information about the future state of the world in one's origin market, risk aversion, and incomplete consumption smoothing lead workers to migrate. In some cases, workers hedge their bets against volatility by migrating to avoid uncertainty.

### 3.2 Foreign firm's problem

Although workers are the primary focus of this paper, firms in the foreign market must be willing to hire laborers under forward contracts in order for migration to occur. I assume that the wage $\tilde{w}$ offered under the contract is set exogenously (e.g., by law) and that laborers across markets are perfect substitutes. Workers are able to pursue an outside option that pays $w_{ft}$.

Several different reasons emerge why a firm would want to hire labor under a forward contract. First, there could exist wage wedges between markets, and forward contracts are an arbitrage opportunity for firms. Second, firms could face problems with labor shortages. Third, human capital acquisition may induce firms to use forward contracts as a form of apprenticeship in order to capture the gains from learning. I briefly elaborate on these reasons below.

There are $N$ homogenous, infinitely lived firms in the foreign market, and they are allowed to import labor and produce using a CRTS function $Q_t = F(K_t, L_t)$, where $K_1$ is fixed.[10] The final good in this sector is sold at a world price $p_t$, and firms are price takers. Capital is available at a constant rental rate $r$. Firms make their decision to import labor under forward contracts and their capital decisions in period 1. Laborers are available to produce in time 1, but new capital is only available in period 2.

Under perfect substitution between laborers, a firm that is able to import labor under forward

[10] A generic CRTS function is not a necessary condition but imposes more stringent restrictions. A Leontieff function or even a production function showing IRTS in $L$ initially, both of which are more appropriate for my historical setting, offer even starker predictions. Alternatively, this function could include land $T$ as a fixed factor.

contracts does so if two conditions hold:

1. $(1+\beta)\tilde{w} \leq w_{f1} + \beta E[w_{f2}]$
2. $(1+\beta)\tilde{w} \leq w_{o1} + \beta E[w_{o2}]$[11]

where $w_{ot}$ is the wage offer that a firm would make to workers if there were no outside option and labor markets were perfectly competitive. The first condition ensures that firms demand outside labor under forward contracts rather than local free laborers, and the second condition ensures that the firm produces and does not shut down.

The wage wedge hypothesis focuses on condition 1 and stems from a large wedge between forward contract wages and free, local wages set by the outside option. Labor shortages in market 1 raise the right hand sides of conditions 1 and 2. Forward contracts create monopsony power and thus lower the arbitrage opportunities for immigrant workers. Firms earn more profits by undercutting laborers from market 1 with laborers from market 0. Compensating differentials may play a role here if the work is onerous enough to raise $w_{ft}$. The statutory wage may not account for this, perhaps intentionally, and so $\tilde{w} \ll w_{ft}$.

If the production function is more labor-intensive for some fraction $\theta \in (0, 1)$ of the total production time,[12] then labor rationing may occur during $\theta$. This creates a hold-up problem for employers in which laborers can attempt to extract higher wages during $\theta$. With mobile labor, some firms could lose actual production while other firms could produce but at the cost of paying out higher spot wages. Following Domar (1970), the use of forward contracts—here, to create a bonded labor force—bypasses labor shortages during critical production periods. For firms, this is analogous to implicit wage contracts with non-existent labor mobility. Here, the institutional constraint on mobility due to indentureship makes the contract acceptable to firms and gives them positive profits.[13]

Finally, forward contracts to capture the gains of specific human capital may exist if $L = g(L, H)$ where $H$ is human capital gained through on-the-job experience or training. With forward contracts, the same firm would bear any training cost (or lower production during training) in return for the higher gains in the future.

[11] Both conditions could include a fixed transportation cost $C$ on the left-hand side, which merely shifts down the indifference value of $\tilde{w}$.

[12] This is true for many agricultural products, including sugar cane, due to a set harvest season.

[13] See Beaudry and DiNardo 1991 for how labor mobility leads to weakly negative profits.

### 3.3 Migration market clearing

A necessary condition for markets to clear and for migration to occur is the statutory wage $\tilde{w}$ must satisfy the worker's and the firm's maximization problems. The exact value of $\tilde{w}$ may vary depending on the decision maker's preference for distributing the surplus. Provided that firms in market 1 are wiling to hire laborers from market 0 under forward contracts, the migrant labor market clears at wage $\tilde{w}$, and a total of $\tilde{L}_t$ workers migrate.[14]

A key aspect of market clearing is the cross-market, cross-time credibility of contracts. Proactive governmental policies (e.g., wage setting, migration facilitation) may be necessary signals of contract credibility that fail in the absence of government action. Because the migrant workers are by definition not local, private contract enforcement may fail if local firms are able to influence the judiciary or comprise the judiciary. For a finite $N$ and particularly as $N \to 1$, this may become more likely. Government participation in migration may be necessary to enforce contracts. Firms benefit from monopsony power and agree to this arrangement, while the arbitrage opportunity for workers encourages them to migrate.

## 4 Data

The existence of demand for indentured servants working under forward contracts is apparent in the over one million indentured servants seen in Table 1. Therefore, I turn to the push and pull factors of local prices and wages in India compared against the ex-ante indentureship contracts to examine why people left India under forward contracts. I focus on Fiji, Jamaica, Natal, and Suriname in order to provide greater coverage of individual records and destinations. I utilize data from several sources to construct both Indian district-level panels and individual-level data.

The primary data source is the emigration pass issued to individuals on departure from India. The emigration pass recorded information such as name, father's name, caste, age, and place of origin were recorded. Due to the bureaucratic nature of the passes, which were created to identify indentured servants both in India and in the destination colony, the data are presumed to represent the population accurately. On the one hand, administrators received no benefit from falsifying records in India. Poor identification of a person could hinder the enforcement of the contract, which

[14] However, as McKenzie et al. (2014) point out, there could be either excess supply or excess demand of laborers.

was adjudicated in criminal court. On the other hand, while there could be some manipulation by the immigrants themselves, who furnished many of the answers, the passes were used to assist with sending remittances, future repatriation, and locating relatives for estate bequests in case of death. Overall, therefore, the data should accurately represent the population leaving India.

I digitized individual-level data from the emigration passes of all 60,000-plus indentured immigrants to Fiji held by the National Library of Australia. I also have digitized repatriation and mortality records through 1915 for the return-migration analysis. I digitized data from the last 12 years of Jamaica's indentureship based on original records held by the National Archives of Jamaica. I obtained various data for Natal and Suriname from the University of KwaZulu-Natal and the Dutch National Archives, respectively.

I entered contract information provided by various years of the *Annual Report of the Protector of Emigrants* in Calcutta, either directly or quoted in Gujadhar (2007). As noted, contracts were set beforehand by law. These reports were created in Calcutta by the Protector of Emigrants, the chief official who oversaw overseas emigration, based on information from India and the recipient colonies. The reports include standardized tables outlining the contracts.

Wages in Fiji, Jamaica, and Suriname were set at one shilling per diem during the entire period. Wages in Natal were set at 10 shillings per month. However, Natal's contracts stipulated complete provision of food for the duration of the contract, while the other colonies provided food for an automatic deduction for a specified period of time.

An immigrant's district of origin links her to district-level prices and wages. I impute missing prices and wages from the nearest non-missing district. Distances are based on district headquarters. For example, Ujjain, a district in the Central Provinces (modern-day Madhya Pradesh), receives prices and wages from Indore, a princely state bordering it to the south. The eponymous district headquarters are roughly 50km from one another. I only observe the characteristics of immigrants and not of all people in a given district. This creates a data set of only 'treated' individuals in which migration is the treatment.

Because I do not observe migration to all colonies with the microdata, I digitized district-year totals of indentured immigrants for available districts from various *Annual Report of the Protector of Emigrants* from Calcutta to compare against my numbers. These are effectively restricted to north India due to record collection. To the best of my knowledge, no comparable data exist for migration from Madras, and state-level reports do not give information on district-specific flows. The data are aggregated across destinations and sometimes across districts in India. For example, Bengal east of Calcutta is often divided into east Bengal and Assam rather than separate districts. Small districts are sometimes combined together, e.g., "Other NWP" captures several thousand people from 1904-1908.

For Bihar and the United Provinces, I have reconstructed approximate totals for main districts prior to 1904. Unfortunately, it is not possible to use the aggregated data from reports in lieu of the microdata due to inconsistencies in the aggregated data. Appendix Table A1 shows that my microdata represent the total data well. Simple cross-tabulations show that over 90% of districts are correctly matched. In regressions of both a binary indicator for any migrants and the total migrants listed in the reports, my respective measures perform well with $r^2$ values of roughly 0.9 and 0.8, respectively. The intensive margin, i.e., the number of migrants reported in the aggregate data appears to be a scalar away from what I construct.

I define a volatility variable similar to prior literature in finance and trade (see, for example, Tenreyro 2007). Volatility for district $j$ in year $t$ is the first difference of the logarithm of annualized prices $p$ over the past $m$ periods:

$$\begin{aligned} z_{j,t,m} &= \ln p_{j,t-m+1} - \ln p_{j,t-m} \\ \nu_{jt} &= SD[z_{j,t,m}], \quad m = 1, \ldots, 5 \end{aligned} \tag{2}$$

I use five years ($m = 5$) in order to match the contract length in the recipient colony.[15] The measure is defined within each district. I focus on rice prices rather than a measure of real wages for both data and theoretical reasons. Wage data were collected in fewer districts than prices and in a coarser fashion than prices. For example, a wage might be designated as "2 to 4 rupees." This provides imprecise information that is not useful to calculate real wages. The magnitude of the wage differential between districts in India and colonies should be correct, but the exact differential (and volatility) may be incorrect. Wages also appear to be sticky with few adjustments over the years, which could result from the data collection. While nominal wages could have moved in the above example between two and four rupees, the reported wage may not have changed. Without

[15] I check for robustness using Chowdhury's (1993) slightly different measure. The correlation between the two measures is almost 1.

more information or stronger assumptions on the distribution of wages within a range over time, rather than my assumption of a symmetric distribution, real wages cannot be calculated correctly.

The method of collecting wages was revised in the early 1900s, whereas prices were collected consistently from 1872 onwards (*Prices and Wages in India* 1921). British colonial officials also expressed their belief that staples' price changes, rather than a fall in wages, explained changes in well-being because of the minimal responsiveness of wages to changing conditions (e.g., *District Gazetteers of the United Provinces of Agra and Oudh* 1904). Wages could have been set more by custom and, in the presence of surplus labor, have been difficult to adjust. Prices, on the other hand, could adjust more easily and clear the markets vis-a-vis a stickier wage.

Using a measure of real wages may impose a functional form assumption on utility. Dividing real wages by a bundle of goods not only presupposes the composition of consumption but also a consumption bundle elasticity of 1. With respect to calories—admittedly not all of consumption—contemporary results in developing countries point to an income elasticity closer to 0 (e.g., Subramanian and Deaton 1996). Subtracting a bundle of goods still requires assumptions on the bundle's composition.

Finally, I focus on rice prices in order to have a consistent commodity across districts. Unlike other crops, rice is reported for almost every district-year pair. The correlation between rice and alternative crops is quite high.[16] Additionally, given rice's wide geographic scope and trade across my study area, Ceylon, and Burma, local, non-rice crops' prices may fluctuate more than rice. Hurd (1975) and Donaldson (2012) find that railroad construction led to price convergence of traded commodities during the same time period as indentureship. This biases me against finding a result for volatility.

The functional form of volatility relies on a view of adaptive expectations in which past realizations of price and volatility affect the expectation of the future outcome. To justify this, I run autoregressive models of log rice price with one lag using my dynamic panel. Despite having 40+ years of data, consistency may be an issue with a static model because prices show autocorrelation.

[16]The correlations between rice and bajra, jawar, ragi, and wheat are 0.74, 0.72, 0.84, and 0.7, respectively. The correlation between rice and wheat, the second most important crop, rises to 0.77 when limiting the sample to the core wheat areas of Central India, the Central Provinces, Punjab and the NWF Province, Rajputana, Sind and Baluchistan, and the United Provinces.

The equations are a simple autoregressive process:

$$y_{jt} = \rho y_{j,t-1} + \gamma_j + \delta_t + \epsilon_{jt} \tag{3}$$

where $j$ indexes the district, $t$ indexes time, and $y$ is the log of rice price. I take the first difference in order to use the Arellano and Bond (1991) GMM estimator. Table 2 shows the results with and without $\delta_t$. For both equations of interest, $\rho$ is positive, significantly higher than 0, and significantly lower than 1, which implies stationarity and autocorrelation. Without time fixed effects, the value of $\rho$ is around 0.89, though this drops to just below 0.7 with fixed effects. Overall, this shows that past price realizations could be used correctly as part of a forecasting model.

# 5 Empirics

I run three main sets of analyses. First, I examine migration from Indian districts to study the impact of volatility on the intensive and extensive margins of migration. Second, I exploit the richness of the microdata to check how volatility changed the composition of the indentured immigrant population. Third, I present results on the long-run impact of initial conditions, including volatility, on return migration to India.

## 5.1 District-level extensive and intensive margins

I aggregate data at the year-district level of departures to examine changes in outflows given price and wage changes. I focus on sending districts that ever sent one migrant. This reduces the problem of differing institutional environments and distances from ports that could differentiate districts that did send migrants from those that did not. Figure 3 shows the geographic distribution of districts that ever sent indentured immigrants to Fiji, Jamaica, Natal, or Suriname. This restriction mostly affects marginal areas in Bengal and Assam in the northeast towards Burma and parts of western and northwestern India; tribal and sparsely populated jungle areas in modern-day Jharkhand and Odisha; and parts of central Deccan. The main sending areas were in the northern Indo-Gangetic plain and along the western coast. I also exclude years in which immigration was suspended, since a zero in such years differs from a zero in years during which immigration was allowed.

Although districts in India were becoming more integrated during this period of time, prices and economic conditions varied across nearby localities. Figure 4 plots rice price volatility for two contiguous districts in the eastern United Provinces (the old Northwest Provinces), Basti and Gorakhpur. These were two of the most densely populated districts in India, and their administrative division, Gorakhpur Division, was more densely populated and more populous than Belgium in the late nineteenth and early twentieth centuries. Gorakhpur's and Basti's respective district headquarters were located roughly 70 km from one another, and they were connected with a rail line as early as 1884.[17] Despite their proximity, prices deviated between the two and led to varying levels of volatility. For example, in the 1890s, rice prices between the two districts diverged and Basti showed higher volatility. This example illustrates both the importance of examining outcomes at a district level and the salience of local markets.[18]

Further evidence for the salience of local markets comes from famine relief in the late 1890s. If observed differences, like those above, merely came from imperfect measurement and potentially uncorrelated errors, then there should be no difference in nearby regions in a famine. Two other contiguous districts in the United Provinces, Banda and Fatehpur, provide evidence that prices indeed measure actual differences. The two district headquarters are located roughly 80km apart. The two districts experienced dearth differently in 1896-7: in Fatehpur, 2.38% of the population (around 16,000) received famine relief, whereas 42.13% of the population in Banda (around 300,000 people) received famine relief. Given that there were clear differences in famine relief, price differences apparently do not arise simply due to measurement error.

Table 3 gives summary statistics of district-level and year-level variables. Roughly half of the district-year observations send at least one indentured migrant. Some districts, such as Basti, sent a migrant in every year; on the other hand, Simla and others only sent migrants in one year. Arcot in the Madras Presidency sent so many migrants and was so well known for its position in indenture that *arkaaṭi* entered Hindi as a term for an indentureship recruiter (McGregor 1993). Conditional on sending at least one migrant in a year, a district sent on average 46 migrants. On average, over 5,200 new indentured servants left India each year for one of these four colonies.

Rice prices and wages appear log normal, and Appendix Figures A1 and A2 show the distribution

[17] I infer this from the construction of Sonepur - Monkepur line 1884, which connects these districts together. Later maps and reports provide confirmation that a rail line existed in 1901 and was *not* built from 1891-1901.

[18] Wheat prices show even greater variability.

of log rice prices and log wages, respectively. Rice prices averaged 3.96 Rs/maund, a measure of roughly 37kg. Rice was cheaper in rice districts, which I define as districts in Assam, Bengal (including Bihar and Orissa), Bombay, Coorg, Hyderabad, Madras, and Mysore. Although rice is my main metric, wheat prices are shown for districts in Berar, Central India (princely states), the Central Provinces, the NW Frontier Province, Punjab, Rajputana, Sind, and the United Provinces.

I separate annual prices and wages into those from sending and non-sending districts. (A non-sending district is a district that sent zero migrants in a given year.) Rice prices are barely higher in districts that do not send than in those that do. However, these differences are small, roughly 0.01 Rs (only a few pies) on average.[19] The variance of rice prices is higher in sending districts. Wheat prices are slightly cheaper in non-sending districts, with an average price roughly one anna less. Roughly the same proportion of districts are predominately wheat-growing (or rice-growing) in the two sets. The major distinction between the two is levels of wages: wages are higher by 1.1 Rs in non-sending districts. Given the similarity in food prices, this translates to potentially higher disposable income in non-sending districts.

Figure 5 shows the average monthly wage in India, weighted by actual departures, compared to the wages in the recipient colonies. I use the rupee-sterling exchange rate given by Goldsmith (1983) and the daily colonial wage defined by law less any food disbursal guaranteed by the contract at government-set prices. All colonies deducted money for food initially, and there is evidence that govenrment-set food prices were inflated (see Knapman 1985), which means the initial take-home wage was a lower bound for subsequent wages. Wages from sending districts in India show a slight increase throughout the study period punctuated by some fluctuations. Wages in the colonies rose to their highest (in rupee terms) at the end of the nineteenth century; the flatness at the end of the series reflects the pegging of the previously silver-denominated rupee to the gold standard and a consistent food price set by the state.

I also include wages in non-sending districts, with each district receiving equal weight. As shown by Table 3, the wages in non-sending districts lie above those in sending districts and appear less volatile. There exists a wedge between wages in the colonies and wages in India. Several years are excluded because immigration was not allowed in those years. For example, an 1898 plague outbreak in Madras in 1898 stopped out-migration from there.

[19] 192 pies = 1 rupee

A key outcome of interest is the sending of a migrant from some district $j$. Given the wide coverage of recruiters, each district can be seen as possessing the opportunity to send a migrant. Thus, there may be key sending-market differences across differences that translated into different migration patterns. I then set up the first estimating equation as a test for the probability that a district sent a migrant in a given year. I estimate first a linear probability model in a fixed effects context:

$$any_{jt} = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(price_{jt}) + \beta_3 \ln(MinColonywage_t) + \beta_4 \nu_{jt} + \Gamma I_j + \epsilon_{jt} \tag{4}$$

where $j$ indexes a district, $t$ indexes year, and $\nu_{jt}$ is a measure of food price volatility in a given year. I include prices in the colonial wage based on the initial mandatory food charges, because this was the take-home wage for workers. I include the minimum wage offered across the four recipient colonies in order to get a conservative bound on the pull factor of wages.

As noted above, I focus primarily on rice in order to ensure a comparable food unit across India. Traditional migration models predict a positive, significant coefficient on recipient-colony wages, i.e., a wage-differential story. The coefficient on $\beta_4$ is the coefficient of interest as part of the risk-aversion/consumption-smoothing story.

Table 4 shows regressions based on Equation 4. Column 1 shows all districts and excludes the volatility measure. Rice price and recipient-colony wages are positive and significant at the 10% and 1% level, respectively, which is consistent with theory. Indian wages are not significant, though. The magnitudes differ for rice price and recipient-colony wages: a rise of 10% in recipient-colony wages raises the probability of that district sending a migrant by 1.9%, while a rise in rice prices of 10% only raises the probability of that district sending a migrant by 0.5%. In column 2, the volatility measure is positive and significant at the 1% level. This is consistent with risk aversion and consumption smoothing: higher levels of price volatility induce more uncertainty, which in turn raises the probability of migration. A rise of volatility by 10% raises the probability of migration occurring from that district by almost 2%. Rice price and recipient-colony wages are still both significant (rice price at 5% and recipient-colony wages at 1%) with similar magnitudes, and Indian wages are not significant.

I add year fixed effects in the latter two columns and drop the minimum colony wage, which is no longer interpretable since it varies at the year and not year-district level. Column 3 provides the analogue to column 1. Rice price is no longer significant but Indian wages are at the 10% level. The sign on Indian wages is negative, which is expected. In column 4, the analogue to column 2, the same result emerges. Rice price and Indian wages have similar magnitudes to Column 3. Volatility is significant at the 5% level. A rise in volatility of 10% raises the probability of a district sending a migrant by 1.4%. Overall, volatility appears to be the most important factor in determining whether or not a district sends a migrant.

Using random effects instead of fixed effects gives similar results. Further, I re-run the specifications for Table 4 using a random-effects probit with as an additional robustness check with different functional form assumptions. Appendix Table A2 gives the results. The results are qualitatively the same even though the coefficients are not directly comparable: volatility is positive and significant at the 1% level in columns 2 and at the 5% level in column 4, which adds in year fixed effects.

Prices, wages, and volatility affect migration not only on the extensive margin but also on the intensive margin. The number of total indentured servants who chose to migrate is also an outcome of interest. An upward-sloping labor supply curve would lead to an increase in migrants in response to higher wages and, in this context, greater uncertainty. Non-convexities in the distribution of the number of migrants may arise due to the mushrooming effect of a first migrant spurring on others to leave.

To address this, I modify the previous equation by using the log of indentured immigrants in each district-year as the dependent variable. The analogue to Equation 4 is

$$\ln(Indentured_{jt}) = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(price_{jt}) + \beta_3 \ln(MinColonywage_t) + \beta_4 \nu_{jt} + \alpha_j + \epsilon_{jt} \tag{5}$$

This removes districts that did not send any migrants in a given year and focuses instead on the intensive effects. Table 5 is the analogue to Table 4 and shows the same groups. As before, columns 1 and 2 exclude year fixed effects and columns 3 and 4 include them.

The log of rice price is positive and significant at the 1% level in the first two columns of Table 5 and at the 5% level in the last two columns. The magnitude is similar across all specifications: a 10% rise in rice prices raises the number of indentured servants by 2.5-3%. Recipient-colony wages are positive and significant at the 1% level in both columns 1 and 2. A rise in these wages raises the number of indentured migrants roughly 1 for 1. Volatility is positive, large, and statistically significant at the 5% level in column 2 but not significant in column 4. The coefficient on volatility is roughly 0.5, or half the coefficient on the recipient colony wage. Log wages in India are correctly signed but not significant in all columns. This result accords with theory and with the binary choice above. Rising prices should lead to higher migration, which is observed. However, the price effect operates immediately through a price shock, not through volatility of prices.

The lack of significance in column 4 and contrast with the extensive-margin results may come from the non-convexity: once one person leaves, many people leave. This could happen because of a mushrooming effect, chain migration, or other similar reason.

To analyze the two dependent variables together and check if the results still hold given the drop in significance from extensive- to intensive-margin regressions, I run Poisson regressions on the counts of indentured servants. The equation is of the form

$$E[Indenture_{jt}|X_{jt}] = \exp(\theta' X) \tag{6}$$

where $X$ is a vector that contains the same covariates as above: log of the minimum recipient-colony wage, log rice price, log wage, rice price volatility, and district fixed effects. Table 6 shows the results from this regression.

The signs of the coefficients match the theoretical expectation in all columns. Recipient-colony wage is significant in columns 1 and 2 at the 1% level. Although volatility is only significant at the 10% level in column 2, it is significant at the 5% level in column 4. Log rice price is significant at the 1% level in all columns. This strengthens the evidence from the previous three tables that volatility matters in out-migration under indentureship contracts.

As a robustness check due to a high number of zeros in the data, I run negative binomial regressions of the same form as the Poisson regressions. Appendix Table A3 shows these results. Volatility is positive and significant in column 2 (without year fixed effects), but it is no longer significant in column 4 (with year fixed effects). However, this is largely consistent with the Poisson results.

As noted above, I could introduce selection by not having the data for other colonies. I check for robustness by replacing my data with the corresponding aggregate numbers for the United Provinces and Bihar. I then re-run Equations 4, 5, and 6 against my main variables of interest and a dummy variable for a replaced year. Appendix Table A4 shows these results. The magnitudes show some attenuation and there is loss of significance for the intensive regressions. However, these are largely robust to this change. Additionally, these numbers are a mix of my own calculations and official data, so there are potential problems mixing numbers that even including a dummy variable cannot fully address.

The results are also robust to dropping the peripheral areas of Assam, Berar, Coorg, and Sind as shown in Appendix Table A5. These provinces were far away from the main ports and were often literally inaccessible.[20] The coefficients change slightly when these areas are dropped but the significance and magnitudes are mostly unchanged. The results are also robust to using a detrended measure of volatility.

Due to the emphasis in the prior literature on wage differentials, I impose a functional form restriction and instead insert the wage differential in place of the local wage and overseas wage. Appendix Table A6 shows the results here. One advantage with this approach is that including year fixed effects no longer forces me to drop overseas wage. However, since some districts have a negative differential, i.e., wages are higher in India than in the colonies, I recode negative differentials to zero and include a dummy variable for this. The results with this additional restriction are consistent with the prior results with no restriction on the relation between domestic and overseas wages.

The above models pool all the colonies together, but there may have been variation in the role of prices in migration to specific colonies. In particular, the contracts for Natal paid lower wages than the other colonies but included food for the entire duration of the contract rather than some fraction. These contractual differences could lead to different take-up across colonies, and the completely guaranteed nature of Natal's contracts predict the greater significance of origin-market volatility for those who went to Natal.

To examine colony-specific migration patterns, I run panel seemingly unrelated regressions on the binary variable that indicates if a district $j$ sent at least one indentured servant to colony $c$.

[20] Sharma (2011)

The model stacks equations from the four colonies of the form

$$any_{jct} = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(price_{jt}) + \beta_3 \ln(wage_{ct}) + \beta_4 \nu_{jt} + \beta_5 exclusion_{jct} + \Gamma I_j + \epsilon_{jct} \quad (7)$$

where $c$ indexes the colony, $j$ the district, and $t$ the year. The error terms are allowed to be correlated across $c$. Dropping a colony-district-year observation leads to dropping the entire district-year observation, so I include an extra dummy variable $exclusion_{jct}$ that indicates that a colony $c$ was excluded from indentureship recruiting in district $j$ during a given year $t$. This is to ensure that $Y$, the matrix of stacked outcomes, is defined across all districts for a given year.[21]

Table 7 gives the results from three SUR models. In column 1, I run Equation 7 as given. The coefficient on log wage in the home district is negative and significant at the 1% level across all four colonies. Wage in the colony is positive and significant at the 1% level except for Jamaica, where it is not significant. Finally, volatility is significant at the 1% level for all four colonies, but it is negative for Fiji and positive for the remainder. The inter-colony differences in volatility and recipient-colony wage from this model stand out. Natal shows the largest magnitudes on its covariates of all the colonies. Fiji's negative coefficient on volatility is puzzling, but the model does not consider spillover effects from other colonies except in the error term.

To deal with the the pull factors of other colonies, I include wages from the other colonies in column 2. The coefficients on volatility for Fiji and Suriname flip, and all volatility coefficients are significant at the 1% level. The preferred specification is column 3, which includes the exclusion dummies for other colonies but not other colonies' wages. This is preferred because columns 1 and 2 assume that other colonies are potential destinations when they are sometimes not. The volatility coefficients for Fiji and Natal, the two biggest colonies, are positive and significant at the 1% level. The volatility coefficients for Jamaica and Suriname are not significant. Own-colony wage is positive and significant at the 1% level in all panels.[22]

Overall, volatility emerges as significant for both the intensive and extensive margins. In general, the covariates are signed correctly according to theory (recipient-colony wages and prices are positive, sending-district wages are negative). The effects persist even when disaggregating to each

[21] I exclude models including a year fixed effect due to computational problems with invertibility.

[22] The logical column 4 including other colony wages and exclusion dummies would be equivalent to four separate OLS regressions per Kruskal's Theorem, since the covariates do not vary between models. I omit running this to use the pooled regressions above.

individual colony.

## 5.2 Composition

I turn to the composition of indentured immigrants using the individual-level microdata. The above section treats immigrants as similar across characteristics and glides over demographic differences over time. Caste and height are the two main individual characteristics of interest. Caste, which here also includes religion, provides more information on the backgrounds of the indentured immigrants. Height serves as a measure of stored nutrition and, in other contexts, is highly correlated with a higher socioeconomic status.

I divide my sample between north and south Indian districts. For caste, this simplifies comparisons and reduces the assumptions required for the comparability across areas. For height, there could exist differences in the population due to differences between these two regions, which are themselves diverse.

### 5.2.1 Caste selection

A different form of selection along caste lines may respond more to volatility and prices than height. Both populations within each broad area (north Indians from Calcutta, south Indians from Madras) show heterogeneity by caste. The inflows of new indentured immigrants varied widely in their composition, and different castes could exhibit different behaviors.

Potential problems arise with using caste as a measure of status. First, caste had become delinked from some occupations even in the time period of study. Second, there is considerable debate over the fluidity of caste and the modern invention of caste as a categorizing tool by British imperial administrators. Third, the contextual nature of caste makes cross-locational comparisons difficult. Localized north Indian castes presumably had no relevance not only in south India but in other parts of north India.

However, I propose to use caste differently and bypass these problems. First, castes were generally associated with a certain socioeconomic status even if the occupational relevance of caste had diminished. In particular, some castes were more likely to own land than others. Second, different castes faced different costs to migrating. In the model's terms, the difference in $A_j$ and $A_f$ could vary by caste, which in turn affects migration decisions. I exploit both intra- and inter-district variation to examine these effects.

I focus here on north Indian castes, which are more easily identified and aggregated in the data. Note that "caste" here refers not only to social subgroups for Hindus but also includes two additional categories for Christians and Muslims. Table 8 gives basic summary statistics for age, height, and several main castes. Note that height is only defined for adults as described above.

The five castes chosen are Muslims plus four Hindu castes plus Muslims. Muslim is a catch-all group that encompasses not only persons explicitly denoted as Muslims but also those from caste groups that were largely Muslim (e.g., Julahas). The Hindu groups are

- **Brahman:** high caste theoretically associated with religious work, though found in large numbers in agriculture and other occupations.
- **Kshatriya:**[23] high caste, traditional landowners
- **Ahir:** middle caste, associated with herding work and non-landowning agriculture
- **Chamar:** low caste, worked with leather and in various laboring positions

As shown by Table 8, these five groups comprised over half of the north Indians who travelled to Fiji, Jamaica, Natal, or Suriname. Selection into migration by different castes could result from differential responses to price and wage shifts, for instance, and the wedge between India and the colonies in terms of amenities.

For comparability with the previous results, I re-run Equations 4, 5, and 6 separately on each of my five main caste groups to examine the extensive and intensive margins of migration along with robustness using a count model (Poisson).[24] For the extensive margin, I model any out-migration by caste as

$$any_{jkt} = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(price_{jt}) + \beta_3 \ln(MinColonywage_t) + \beta_4 \nu_{jt} + \Gamma_1 I_j + \Gamma_2 I_t + \epsilon_{jt} \quad (8)$$

where the $k$ subscript now indexes the caste. Table 9 shows the results for these regressions. In contrast to Table 4, both local (Indian) prices and wages matter for several castes. A rise in rice prices is significantly associated with out-migration by Brahmans (5% level), Ahirs (10% level),

[23]Here, Kshatriya also includes Jats, Rajputs, and Thakurs.

[24]This is the same as running SUR with the five caste groups separately without model-varying covariates.

and Chamars (5% level). Similarly, for these same castes, a rise in wages is associated with lower out-migration (all significant at the 1% level).

Volatility matters for only one caste at the extensive margin: Kshatriyas. A 10% rise in volatility leads to roughly 1.8% increase in probability of Kshatriya out-migration, and this is significant at the 5% level. The magnitude of this coefficient is roughly the same as in Table 4. This appears puzzling at first: Kshatriyas were largely landowners and were noted to be a caste of greater economic means. However, the lack of significance and small magnitudes for the other castes does not constitute evidence for their ability to smooth. Instead, it could also be that the other castes were unable to smooth consumption at *any* level. Kshatriyas, in contrast, show a significant positive gradient consistent with a decreasing ability to smooth consumption at higher levels of volatility.

I next run intensive-margin regressions in which I replace the binary dependent indicator for any member of a caste with the log of caste migration. Table 10 shows these regressions. The results are largely the same as the previous table. For Brahmans, Ahirs, Chamars, and Muslims—but not Kshatriyas—a 10% increase in rice prices increased the number of out-migrants by those castes from 4%-7%. For all four, this is significant at the 1% level. In contrast, the sign is negative but not significant for Kshatriyas. For wages, Ahir, Chamar, and Muslim out-migrants drop by 4.5%-8.7% for a 10% increase in wages. Brahmans also show a negative though insignificant coefficient, but Kshatriyas show a positive but insignificant coefficient. Volatility is not significant for any caste.

Finally, I run Poisson regressions because the underlying data are counts. Table 11 shows these regressions. The results are qualitatively similar, but now only rice price is significant and only for Ahirs, Chamars, and Muslims. As with the previous two tables, high rice prices are associated with higher out-migration. Brahmans, too, show the same positive sign but insignificant coefficient. Kshatriyas, though, have a negative but insignificant coefficient.

Overall, the previous three tables lend support to three conclusions at the aggregate level. First, while I do not test for coefficient equality across castes, different castes responded differently to local economic conditions. Second, the division appears to come from landowners (Kshatriyas) and non-landowners. The weaker and sometimes insignificant results from Muslims may reflect problems with pooling this heterogeneous group. Some Muslim groups, such as Sheikhs and Syeds, did own large amounts of land, and their inclusion may cause the attenuation that I observe. Third, the role of volatility is ambiguous. On the one hand, Kshatriyas on the extensive margin do show an increase in migration in response to higher volatility, which matches both the model and the earlier pooled regressions. On the other hand, the other groups that a priori should have been affected do not seem to be. I attribute this to the other groups' inability to smooth rather than a perfect ability to smooth.

I also run regressions for each of these caste groups pooled and then separately by sex/age: male adults, female adults, and children of both sexes. Appendix Table A7 shows the results on the extensive margin for male adults, female adults, and children separately. The results looks similar to the extensive margin pooled across sex and age. Appendix Tables A8 - A10 show the results on the intensive margin for male adults, female adults, and children. These tables largely resemble Table 10 except that the coefficient for rice price for Chamars grows in magnitude from males to females and children. This could reflect different migration decisions for families rather than for males, many of whom came singly.

I now turn to the individual-level data to examine how the share of each caste changed as economic conditions changed in India. Not only do the incidence and levels matter, but the relative size of each group also speaks to a caste's ability to deal with prices, wages, and volatility. I separate male adults, female adults, and children like above in order for an analysis on adults to not pick up an effect of, for example, one caste bearing more children and/or bringing more family members. I run the following logit equation:

$$\begin{aligned}(caste_{ijt} = k | sex_{ijt} = g) = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(price_{jt}) + \beta_3 \ln(Colonywage_{ct}) \\ + \beta_4 \nu_{jt} + \Gamma_1 D_t + \Gamma_2 D_j + \Gamma_3 D_c + \epsilon_{ijt}\end{aligned} \tag{9}$$

where $i$ indexes person, $j$ indexes district, $c$ indexes colony, and $t$ indexes year. $g$, the categories for $sex$, takes on three different values: man, woman, or child. $Caste$ is a categorical variable containing the five castes mentioned above plus an omitted caste category, and $k$ gives the different values of $caste$. $\nu_{jt}$ is the same volatility variable as used above.

Table 12 provides the results for each caste one by one. The even columns contain the year fixed effects, which mute some of the effects. Three main points stand out. First, the draw of higher colony wages pull in proportionally more Brahman and Muslim males. Second, across all three groups, rice price continues to act for Kshatriyas opposite to everyone else. Higher rice prices are associated with significantly lower shares of Kshatriyas (significant at the 1% level for male adults, female adults, and children). Third, the proportions of Chamars for adult males, adult females, and children significantly increase with rises in rice. The female coefficient is triple double that of the male coefficient and the children's coefficient is five times that of the male coefficient. In part, this reflects a larger sample of male adults so that the effects are not as drastic. However, it does indicate that Chamar families may have been more responsive than other families to rises in rice prices.

I re-run Equation 9 as a multinomial logit with Chamar as the omitted category. The interpretation of the coefficients is the impact on the share of a particular caste within the indentureship population relative to Chamars. This is important given the lower status of Chamars and their status as laborers, split roughly evenly between agricultural and non-agricultural work. Appendix Table A11 shows the results from these regressions. Chamar is the omitted category. Even columns include year fixed effects and are the main columns of analysis. I pool children of both sexes together. The results are qualitatively similar to the one-by-one logits. Interestingly, for all three age*sex groups, no caste share appears different from Chamars as volatility changes. The Kshatriya coefficient is positive and somewhat large but insignificant.

Overall, the composition of indentured servants changes in response to prices, wages, and volatility. Castes show different responsiveness to the pull factor, recipient-colony wages. Additionally, there is evidence that castes exhibit different abilities to smooth consumption. In the face of immediate changes in Indian wages and especially prices, Chamars increased their shares but not evenly. There is some evidence that Kshatriyas were able to smooth consumption in the face of higher volatility but their share did not change with an increase of volatility. This result gives suggestive evidence of caste-specific insurance markets roughly 100 years before Munshi and Rosenzweig's (2016) data.

### 5.2.2 Height

Height serves as a marker of stored health and socioeconomic status (see Bleakley et al. 2013 for an overview). Migrants may be selected on height (e.g., Kosack and Ward 2014, Spitzer and Zimran 2014). At the same time, measured heights can drastically change due to selection (see, e.g., Zimran 2015). Height differentiations over time in the sending population give evidence on the form of selection. The role of volatility in selection along height is ambiguous, though. Adult height is already pre-determined and is not a choice, and volatility may equally affect adults of varying heights. I examine what forms of selection occur based on height compared to prices.

Appendix Figures A3 and A4 show kernel densities for the heights of males and females, respectively, overall and by port of departure. I censor the lower bound of heights at four feet. Port of departure roughly corresponds with a north-south divide, with north Indians going via Calcutta and south Indians via Madras. The pooled and the separate distributions appear roughly normal. In both cases, as seen also in the summary statistics, immigrants from Madras are taller than their Calcutta counterparts.

Table 13 summarizes the differences in heights for both men and women. As seen in Figure A3, males from both ports appear to have similar heights. After controlling for year of departure and age at departure, Madras men were, on average, just as tall than Calcutta men (column 4). On the other hand, Madras females were around 0.622 inches (1.6) cm taller than Calcutta females (column 8), and this is significant at the 1% level.

Because of the differences in mean heights and in geography between the two groups, I analyze them both together and separately by port. I estimate the following equation for adult height separately by sex:

$$\begin{aligned}\ln(Height_{ijt}) = \beta_0 + \beta_1 \ln(wage_{jt}) + \beta_2 \ln(riceprice_{jt}) + \beta_3 \ln(Colonywage_{ct}) + \beta_4 \nu_{jt} \\ + \beta_5 I_{Madras} + \beta_6 Age + \beta_7 Age^2 + \Gamma_1 D_t + \Gamma_2 D_j + \Gamma_3 D_c + \epsilon_{ijt}\end{aligned} \tag{10}$$

where $i$ indexes a person, $j$ indexes a district, $c$ indexes a colony, and $t$ indexes a year. $D_t$, $D_j$, and $D_c$ are vectors of year, district, and recipient colony fixed effects, respectively. $\nu_{jt}$ is price volatility as described above. I include the dummy variable for Madras only in pooled specifications. I cluster standard errors at the district level.

Table 14 shows the results for men. In odd columns I exclude the volatility term. Age and age squared are significant at the 1% level in all specifications and have opposite signs. This makes sense: older people tend to be taller, but the advantage to age diminishes over time because people stop growing. Columns 1 and 2 pool individuals from both north and south. In columns 1-4, log rice price is negative and significant at the 1% level; this effect in the pooled columns 1-2 comes entirely from Calcutta migrants. The magnitudes are small, though. Log wage is positive in all specifications, though its significance drops from 5% in columns 1-4 to 10% in columns 5-6. Wages in the recipient colony show more variation across port. 10% higher recipient-colony wages are associated with a drop in height of 0.07%, almost half an inch off the mean, in Calcutta, and this is significant at the 1% level. 10% higher recipient-colony wages are associated with an increase in height of almost 0.9% from Madras, and this is significant at the 1% level.

As wages in their home districts decreased or rice prices increased, shorter men migrated. Under the assumption of a positive correlation between socioeconomic outcomes and height, this points to negative selection amongst the male indentured immigrants. However, selection based on recipient-colony wages differed across port, with Madras showing positive selection and Calcutta showing more negative selection.

Females exhibit the same age pattern as males but different price trends. Table 15 shows the female regressions. Like males, age and age squared are significant at the 1% level in all columns However, log rice price is negative and significant at the 5% level just for Madras (columns 5 and 6). Like with men, the magnitudes are small: a fall in wages of 10% raises female height by 0.28%. This corresponds to a drop in 0.17 inches (4.1 mm). The volatility term is not significant in either the pooled regression or the port-by-port regressions. This is consistent with negative selection amongst the Madras female indentured immigrants.

In all cases except the pooled regressions for females, the volatility term is not significant, but this is unsurprising. First, I measure volatility based on outcomes over the five years prior to migration. Economic conditions early in life for migrants, sometimes 20-30 years prior to immigration, may have had little correlation to volatility just before migrating. Second, even if volatility of outcomes were highly correlated across decades, gains from good periods could, on average, offset losses from bad ones if the good ones occurred during critical growth years.

### 5.2.3 Return Migration

The initial conditions at migration may exert a long-run effect on behavior. Migrants who left under worse conditions may choose to stay rather than return home. Put differently, worse conditions at home lowered the cost for permanent migrants to leave in the first place. Furthermore, spending time in the colony abroad resolved uncertainty about the local conditions and is the reverse of the original model, i.e., migrating now back to uncertainty in India and away from certainty in the colonies.

A novel element of the data for Fiji is the tracking of individuals after their arrival due to the contractual offer of a return passage to India after 10 years in the colony, of which at least five were required to be served as an indentured laborer. This meant officials kept extensive records of deaths and actual departures from the colony. The National Library of Australia holds mortality records from the start of indentureship through 1927 and return migration records up through the last repatriations in the 1950s. Therefore, unlike other work on return migration (e.g., Abramitzky et al. 2012), I am able to discover the fate—stayed, died, or returned—of each individual at any given point in time after migration.

I focus on the cohorts who entered Fiji in 1903 or earlier. Workers who entered in 1904 or later were affected by World War I, which reduced the availability of transportation back to India. As noted, Indians were contractually guaranteed free passage back to India after ten years. However, some indentured immigrants were repatriated in their first five years due to disease or physical inability to work. Such repatriation was subject to a medical officer's decision and not a choice. Others chose to pay their own passage back after indentureship but before serving out the full ten years. Who was repatriated due to frailty, who chose to pay, who chose the free ticket, and who chose to remain are salient questions not only in this context but also in general in return migration. Additionally, the interactions between repatriation and mortality are important to study.

This context provides excellent data and institutional details to bypass other problems with studying return migration. First, I bypass issues with migrants being unable to return migrate due to the high cost (e.g., destitute migrants or those who must pay back their initial debts), which was important during that period worldwide and affected migration (see, e.g., Armstrong and Lewis 2012). Second, time preferences matter only after year five. Voluntary return migration only occurred after year five, which forced all indentured immigrants to learn their types and about the local economy in Fiji. If migrants faced initially low free-market wages due to the steep learning curve in the economy, this enforced period of servitude enabled them to learn, earn, and make an informed decision about return migration. Third, I account for mortality and reduce measurement error in attributing return migration to mortality or vice versa.

Table 16 gives summary statistics of the post-indentureship fates of the adult immigrants who arrived in 1903 or earlier. I set the cutoff date to 31 December 1913 in order to not contaminate the analysis with World War I. Due to the later arrival of many migrants from Madras, I split up the two ports. I only report castes for departures from Calcutta. Roughly one third of all Calcutta immigrants returned to India, with an average duration in Fiji of roughly nine years. An additionally 18% perished in Fiji, leaving just under half remaining in the colony. For Madras immigrants, roughly one quarter returned to India and one tenth died in Fiji.

For Calcutta immigrants, different patterns appear across sex and caste. A slightly higher proportion of male adults returned than female adults, while a lower proportion of male adults died than female adults. The cumulative effect of mortality and repatriation, though, roughly equalizes the proportion of male and female adults who remained in Fiji. Children were repatriated at roughly the same rate as female adults, which could reflect children accompanying their mothers, and died at higher rates than adults.

Upper castes (Brahmans and Kshatriyas) returned at higher rates than Chamars, the primary low caste in this analysis.[25] Muslims, though, were also roughly as likely to return. Mortality patterns across castes appear similar but Brahmans show the highest rate while Muslims show the lowest rate. Over half of Chamars introduced before 1903 remained in the colony, while under 30% of Brahmans remained.

Interestingly, Calcutta immigrants returned on average after 9.11 years before receiving their free passages. Additionally, Calcutta immigrants died on average after roughly 5.7 years in the colony, i.e., after the completion of the five-year indentureship project. The comparable numbers for Madras immigrants are lower, which reflects the lower amount of time between arrival and the current cutoff date.

I model the survival of individuals in Fiji with three separate failures: repatriation, mortality, or either. Duration is measured in days from the date of arrival, which was recorded by the National Library of Australia, to death, repatriation, or the end of the data series. I use Cox proportional

[25] See Persaud (2017) for more on caste-specific return migration that exploits the timing and differences in return prices. Return migration here focuses only on the role of volatility.

hazards with time-invariant covariates where the hazard function takes the form

$$\begin{aligned}\lambda(t|X_{ijt}) =& \lambda_0(t)\exp(\beta_1 female_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 height_i + \beta_5 height_i^2 \\ &+ \beta_6 \ln(RicePrice_{jt}) + \beta_7 \ln(wage_{jt}) + \beta_8 \nu_{jt} + \Gamma_1 I_k)\end{aligned} \tag{11}$$

where $i$ indexes a person, $k$ indexes caste, $j$ indexes district, and $t$ indexes the year of departure. $I_k$ is a vector of a larger group of castes. I use a richer set of castes than above per Lal (2004) and McGregor (1993) but report only coefficients on Ahirs, Brahmans, Kshatriyas, and Muslims. Chamars are again the omitted category.[26]

Table 17 gives the results for repatriation. I run different duration models for different periods of repatriation: ever, within the first five years, after five years, in years 5-10, and after year 10. Brahmans and Kshatriyas are always more likely than Chamars to be repatriated, even in the first five years; this indicates poorer health on their part. Muslims are more likely overall, but this is driven by being more likely after five years, not before. Muslims are also more likely to be repatriated in the period of free passage than before, which differs than Brahmans and Kshatriyas. Finally, Ahirs are more likely to be repatriated but this comes from the pay-for-yourself period of years 5-10 rather than later or before. Overall, the results are consistent with upper castes having a higher value of returning to India, which is a gross measure of higher returns, higher amenities, and potentially lower demands from their caste-insurance group.

Women are less likely to return to India than men; being a women reduces the hazard by 25%. This could have resulted from the adult sex imbalance in Fiji that was heavily skewed towards men. Height is negative significant (albeit only at the 10% level) only for years 5-10; height squared, though, is positive and significant, which indicates a diminishing reduction in repatriation probability. Higher food prices in the year of departure are positive and significant in columns 3 and 5 at the 1% level. Given that I control for caste, this could indicate that people who were unable to smooth consumption due to a price shock are more likely to return home.

Table 18 shows similar specifications for mortality. In the first five years, the period of indentureship, only Muslims show a significantly lower mortality rate than Chamars. This is a striking

[26] The full list is Ahir, Brahman, Gararia, Kahar, Kori, Kurmi, Pasi, Lodh, Kewat, Lonia, Murao, Bania, Barhai, Dosad, Bhar, Kachi, Gond, Gowala, Teli, Kumhar, Lohar, Mosaher, Kunbi, Dhobi, Sweeper, Unspecified, Muslim, Kshatriya, and Chamar. In tables, I omit reporting the coefficients except for the main castes listed above in Section 5.2.1.

result because it provides some evidence that flattening the labor market through a statutory wage also flattens the mortality distribution. After five years, Brahmans are over 40% less likely to die than Chamars but this is significant only at the 1% level. Women are significantly more likely to die than men (an increase in the hazard of 17%), but high mortality in the first five years drives this.

Appendix Table A12 uses exit as the duration variable. Exit is defined as either death or repatriation. This table largely shows similar results to Table 17. Women are still less likely to leave, either by repatriation or death, than men, but the coefficient has dropped. towards zero.

The previous tables do not show the dynamic choice of whether or not to return to India, though. I thus model survival using duration in Fiji in years and time-varying covariates. Table 19 shows this model, which includes individuals who are still in Fiji in any given year. In columns 4 and 5, I include both current home-district conditions and initial, anchoring conditions in the year of departure.

As migrants age, they are more likely to return; this appears in all five columns. As noted above, females are less likely to return to India, and the difference in hazards (a drop of roughly 0.2 in all columns) is similar to that in Table 17. In column 3, I add castes. Brahmans, Kshatriyas, and Muslims are significantly more likely than Chamars to return to India, with increased hazards ranging from 0.18 to over 0.4.

In column 4, prices both now and at the time of departure are significant and move largely in the theoretical direction. However, anchor (initial) Fiji wages and current Fiji wages are both positive; this contrasts with the initial migration decision because this means that higher wages push Indians to return to India rather than stay. Current wage is the value of the indentureship wage because full wage data are not available for Fiji. However, evidence exists that point to statutory wages lying below the market wage. Thus, the estimated coefficient for wage provides a lower bound, which means an even stronger impact of wages on return migration.

Given the wage rigidity during indentureship, few adjustment costs due to food, and relatively stable exchange rate, this is consistent with target-income migration. Indians may have returned after earning a certain amount in Fiji; higher wages would then be consistent with higher out-migration. Volatility in the current period has a much larger coefficient than volatility in the anchor period, but both are negative. This is consistent with several competing, though not mutually exclusive, explanations: maintaining negative beliefs about return migration, knowledge of home conditions, and choosing to remain in Fiji in order to not participate in home insurance networks. Column 5 combines both caste and prices and wages, and the results are largely similar to the previous column.

Overall, these results point to two similar potential explanations. First, Indians who came from more volatile regions in the first place were permanent migrants and did not want to return to economically volatility. Second, Indians making decisions in a given period about returning home avoided returning in more economically volatile times even in the same district. This is the mirror of my results in Section 5.1. In the initial selection, Indians fled volatility; in the return selection, Indians chose not to return to avoid volatility.

## 6 Conclusion

Uncertainty and the second moment of consumption are important determinants of migration, particularly when uncertainty can be reduced under a forward contract. The institutional context of Indian indentureship allows me to peg uncertainty on one side of the migration decision and test the impact of prices, wages, and price volatility on migration. I provide new evidence on an unexplored topic, South-South migration in the so-called 'Age of Migration' prior to World War I.

The results show that volatility does matter: higher volatility significantly increases migration from a district on both the extensive and intensive margins. This is consistent with basic economic theory, but also points out areas for public policy interventions in countries that send large numbers of international migrants. Staple-commodity prices and insurance may reduce out-migration, for example. In India, NREGA guarantees an income stream and may dampen demand for guaranteed labor contracts abroad. The historical counterfactual that I study shows that an insurance scheme like NREGA would have responded directly to the needs of out-migrants.

The composition of indentured immigrants bears out the general result but also differential responses to potential policies. Overall height responds more to prices and wages, which is consistent with volatility affecting people of different heights similarly. However, volatility does play a role in caste selection. These results are consistent with a lack of consumption smoothing and the use of migration to overcome negative consumption shocks and long-run uncertainty. Differential responses to prices by caste groups indicate different economic roles—notably, the position of Kshatriyas as landlords—but also different access to insurance. Social networks and intra-group insurance may offer support for some groups but not others, and this appears linked to underlying socioeconomic differences across groups. Public policies that do not take into account group identity, the initial distribution of resources, and intra-group insurance may not effectively target those most vulnerable to volatility.

Finally, return migration data show volatility affected return migration. This paper has implications for modern-day guest worker programs, too, which have been promoted as an alternative to illegal immigration or even legal, permanent migration. However, as shown, some bonded laborers may forego a monetarily costless trip home after their tenures end. Guest workers will re-optimize and not return to economically volatile home markets. Higher volatility may select permanent migrants rather than temporary migrants. Further research on return migration will help answer whether the decision to migrate was part of a larger process of migration, in which a price shock was the proverbial last straw, or if migration was a long, costly method of insurance. Further research will also address intergenerational outcomes to examine an even longer-run effect of volatility at the time of migration. Persaud (2017) examines the composition of return migrants and the economics of identity underlying high upper-caste return migration.

Overall, this paper points to the salience of the second moment, measured here as volatility, in pushing people out of a low-income location into long-term bonded-labor contracts. For takers, the certainty gains from these contracts outweighed the freedom and mobility costs from servitude. Given the persistence of these types of contracts and rising volatility in the world, bonded labor may become more prevalent in the absence of other risk-mitigation mechanisms.

## References


[1] Abramitzky, Ran, and Fabio Braggion. 2006. "Migration and human capital: Self selection of indentured servants to the Americas." *Journal of Economic History* 66 (4): 882-905.

[2] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5): 1832-1856.

[3] Acemoglu, Daron, and Alexander Wolitzky. 2011. "The Economics of Labor Coercion". *Econometrica* 79 (2): 555-600.

[4] Adamson, Alan H. 1972. *Sugar without slaves: the political economy of British Guiana, 1838-1904*. New Haven: Yale University Press.

[5] Arellano, Manuel, and Stephen R. Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58 (194): 277-297.

[6] Armstrong, Alex, and Frank D. Lewis. 2012. "International migration with capital constraints: interpreting migration from the Netherlands to Canada in the 1920s." *Canadian Journal of Economics/Revue Canadienne D'économique* 45 (2): 732-754.

[7] Baker, Scott, Nicholas Bloom, and Steven Davis. "Mesauring Economic Policy Uncertainty." NBER WP 21633.

[8] Beaudry, Paul, and John DiNardo. 1991. "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data." *Journal of Political Economy* 99 (4): 665-688.

[9] Bleakley, Hoyt, Dora L. Costa, and Adriana Lleras-Muney. 2013. Trends in health, education and income in the United States, 1820-2000. NBER WP 19162.

[10] Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3): 623-685.

[11] Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review.* 77 (4): 531-553.

[12] Brennan, Lance, John McDonald, and Ralph Shlomowitz. 1994. "Trends in the Economic Well-Being of South Indians under British Rule: The Anthropometric Evidence." *Explorations in Economic History.* 31 (2): 225-260.

[13] Brennan, Lance, John McDonald, and Ralph Shlomowitz. 1998. "The geographic and social origins of Indian indentured labourers in Mauritius, Natal, Fiji, Guyana and Jamaica." *South Asia: Journal of South Asian Studies.* 21 (sup1): 39-71.

[14] Brennan, Lance, John McDonald, and Ralph Shlomowitz. 2003. "Long-Term Change in Indian Health." *South Asia: Journal of South Asian Studies.* 26 (1): 51-70.

[15] Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica.* 82 (5): 1671-1748.

[16] Deerr, Noel. 1949. *The history of sugar.* London: Chapman and Hall.

[17] Dippel, Christian, Avner Greif, and Daniel Trefler. 2015. "The rents from trade and coercive institutions: removing the sugar coating." NBER working paper 20958.

[18] *District Gazetteers of the United Provinces of Agra and Oudh.* 1904. Allahabad: Government Press.

[19] Domar, Evsey. 1970. "The Causes of Slavery or Serfdom: A Hypothesis." *The Journal of Economic History.* 30 (1): 18-32.

[20] Donaldson, Dave. 2012. "Railroads of the Raj: estimating the impact of transportation infrastructure."

[21] Galenson, David W. 1981. "The Market Evaluation of Human Capital: The Case of Indentured Servitude." *Journal of Political Economy* 89 (3): 446-467.

[22] Galenson, David W. 1984. "The Rise and Fall of Indentured Servitude in the Americas: An Economic Analysis." *The Journal of Economic History* 44 (1): 1-26.

[23] Genicot, Garance. 2002. "Bonded labor and serfdom: a paradox of voluntary choice." *Journal of Development Economics* 67 (1): 101-127.

[24] Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *The American Economic Review.* 60 (1): 126-142.

[25] Honekopp, Elmar. 1997. "Labour Migration to Germany from Central and Eastern Europe - Old and New Trends." Working paper.

[26] Hurd, John. 1975. "Railways and the expansion of markets in India, 1861-1921." *Explorations in Economic History* 12 (3): 263-288.

[27] Jayachandran, Seema. 2006 "Selling labor low: wage responses to productivity shocks in developing countries." *Journal of Political Economy.* 114 (3): 538-575.

[28] Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1): 211-251.

[29] Knapman, Bruce. "The Rise and Fall of the White Sugar Planter in Fiji 1880-1925." *Pacific Studies.* 9 (1): 53-82.

[30] Kosack, Edward and Zachary Ward. "2014. "Who crossed the border? Self-selection of Mexican migrants in the early twentieth century." *Journal of Economic History* 74 (4): 1015-1044.

[31] Lal, Brij V. 2004. *Girmitiyas: the origins of the Fiji Indians.* Lautoka [Fiji]: Fiji Institute of Applied Studies.

[32] Mangru, Basdeo. 1987. *Benevolent neutrality: Indian government policy and labour migration to British Guiana, 1854-1884.* London: Hansib Pub. Ltd.

[33] Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1993. "Theories of International Migration: A Review and Appraisal." *Population and Development Review* 19 (3): 431-466.

[34] McGregor, R. S., ed. 1993. *The Oxford Hindi-English Dictionary.* Oxford: Oxford University Press.

[35] McKenzie, David, Caroline Theoharides, and Dean Yang. 2014. "Distortions in the International Migrant Labor Market: Evidence from Filipino Migration and Wage Responses to

Destination Country Economic Shocks?" *American Economic Journal: Applied Economics*. 6 (2): 49-75.

[36] Munshi, Kaivan, and Mark Rosenzweig. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap." *American Economic Review*. 106 (01): 46-98.

[37] Naidu, Suresh, and Noam Yuchtman. 2013. "Coercive Contract Enforcement: Law and the Labor Market in Nineteenth Century Industrial Britain." *American Economic Review*. 103 (1): 107-144.

[38] Naidu, Suresh, Yaw Nyarko, and Shing-Yi Wang. "Monopsony Power in Migrant Labor Markets: Evidence from the United Arab Emirates." Forthcoming in *Journal of Political Economy*.

[39] Persaud, Alexander. 2017. "A (Paid) Passage to India:
Quantifying the value of upper-caste status." Working paper.

[40] *Prices and Wages in India*. Multiple years. Calcutta: Department of Statistics.

[41] *Report of the immigration agent general*. Multiple years. Georgetown, Guyana: Argosy.

[42] *Report of the Protector of Indian Immigrants*. Multiple years. Pietermaritzburg, South Africa: "Times" Printing and Publishing Company, Ltd.

[43] Persaud, Alexander. 2017. "A (Paid) Passage to India:
Quantifying the value of upper-caste status." Working paper.

[44] Roberts, G. W. 1955. "Emigration from the island of Barbados." *Social and Economic Studies* 4 (3): 245-288.

[45] Rodney, Walter. 1981. *A history of the Guyanese working people, 1881-1905.* Baltimore: Johns Hopkins University Press.

[46] Roopnarine, Lomarsh. 2010. "Re-Indenture, Repatriation and Remittances of Ex-Indentured Indians from Danish St Croix to British India 1863-18731." *Scandinavian Journal of History*. 35 (3): 247-267.

[47] Rosenzweig, Mark R., and Oded Stark. 1989. "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy*. 97 (4): 905-926.

[48] Roy, Tirthankar, and Anand V. Swamy. 2016. *Law and the economy in colonial India.* Chicago: University of Chicago Press.

[49] Sarup, Leela Gujadhur. 2006. *Colonial emigration, 19th, 20th centuries: Annual reports from the port of Calcutta to the British & foreign colonies.* Kolkata: Aldrich International.

[50] Seecharan, Clem. 1997. *'Tiger in the stars:' the anatomy of Indian achievement in British Guiana 1919-1929.* London: Macmillan Education.

[51] Sharma, Jayeeta. 2011. *Empire's garden: Assam and the making of India.* Durham [N.C.]: Duke University Press.

[52] Spitzer, Yannay and Ariell Zimran. 2015. "Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907-1925." Working Paper.

[53] *Statistical Abstract Relating to British India, from 1867/8 to 1876/7.* 1878. London: HM Stationery Office.

[54] Stiglitz, Joseph E. 1974. "Incentives and Risk Sharing in Sharecropping." *The Review of Economic Studies* 41 (2): 219-255.

[55] Subramanian, Shankar, and Angus Deaton. 1996. "The Demand for Food and Calories". *Journal of Political Economy.* 104 (1): 133-162.

[56] Tenreyro, Silvana. 2007. "On the trade impact of nominal exchange rate volatility." *Journal of Development Economics* 82 (2): 485-508.

[57] Tinker, Hugh. 1974. *A New system of slavery: the export of Indian labour overseas, 1830-1920.* London: Oxford University Press.

[58] Turnovsky, Stephen, Haim Shalit, and Andrew Schmitz. 1980. "Consumer's Surplus, Price Instability, and Consumer Welfare." *Econometrica* 48 (1): 135-152.

[59] Yang, Dean. 2006. "Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants' Responses to Exchange Rate Shocks." *Review of Economics and Statistics* 88(4): 715-735.

[60] Zimran, Ariell. 2015. "Does Sample-Selection Bias Explain the Antebellum Puzzle? Evidence from Military Enlistment in the Nineteenth-Century United States." Working paper.

Figure I.1: Price volatility and emigration under indentureship

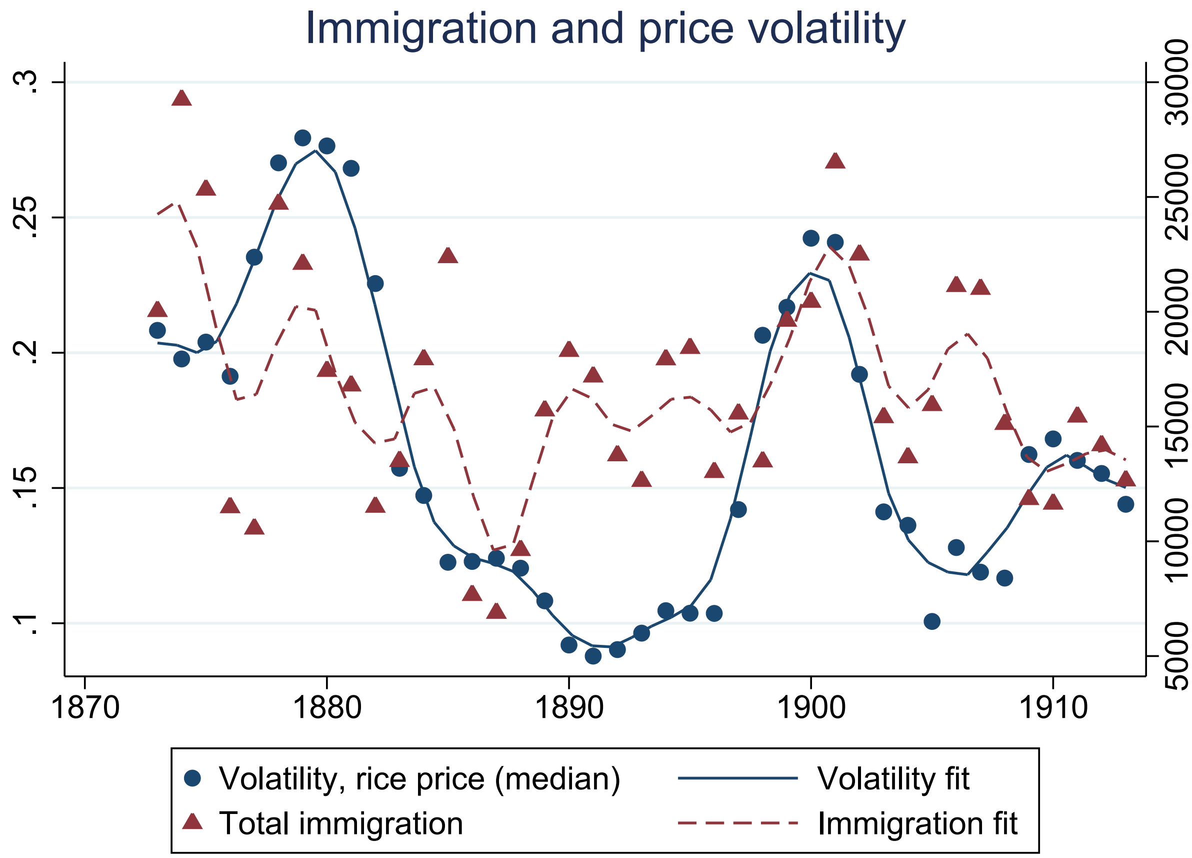
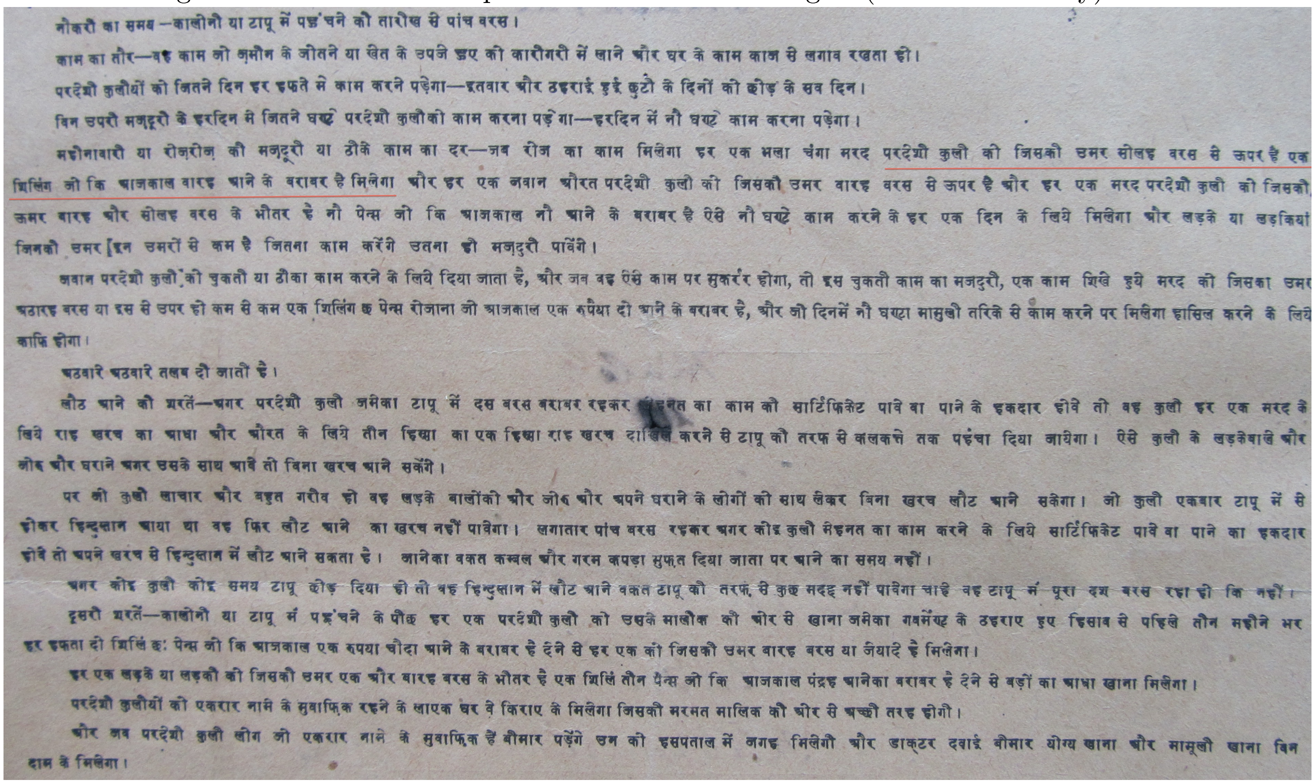


Figure I.2: Contract excerpt for Jamaica in Devanagari (late 19th century)



नौकरी का समय—कालोनी या टापू में पहुंचने की तारीख से पांच बरस।

काम का तौर—वह काम जो ज़मीन के जोतने या खेत के उपजे हुए की कारीगरी में लाने और घर के काम काज से लगाव रखता हो।

परदेशी कुलीयों को जितने दिन हर हफ्ते मे काम करने पड़ेगा—इतवार और ठहराई हुई छुटी के दिनों को छोड़ के सब दिन।

बिन उपरी मजदूरी के हरदिन मे जितने घण्टे परदेशी कुलीको काम करना पड़ेगा—हरदिन में नौ घण्टे काम करना पड़ेगा।

महीनावारी या रोजरोज़ की मजदूरी या ठीके काम का दर—जब रोज का काम मिलेगा हर एक भला चंगा मरद परदेशी कुली को जिसकी उमर सोलह बरस से ऊपर है एक शिलिंग जो कि आजकाल बारह आने के बराबर है मिलेगा और हर एक जवान औरत परदेशी कुली को जिसकी उमर बारह बरस से ऊपर है और हर एक मरद परदेशी कुली को जिसकी ऊमर बारह और सोलह बरस के भीतर है नौ पेन्स जो कि आजकाल नौ आने के बराबर है ऐसे नौ घण्टे काम करने के हर एक दिन के लिये मिलेगा और लड़के या लड़कियां जिनकी उमर [इन उमरों से कम है जितना काम करेंगे उतना ही मजदुरी पावेंगे।

जवान परदेशी कुली को चुकती या ठीका काम करने के लिये दिया जाता है, और जब वह ऐसे काम पर मुकर्रर होगा, तो इस चुकती काम का मजदुरी, एक काम शिखे हुये मरद की जिसका उमर अठारह बरस या इस से उपर हो कम से कम एक शिलिंग छ पेन्स रोजाना जो आजकाल एक रुपैया दो आने के बराबर है, और जो दिनमें नौ घण्टा मामुली तरिके से काम करने पर मिलेगा हासिल करने के लिये काफि होगा।

अठवारे अठवारे तलब दी जाती है।

लौट आने की शरतें—अगर परदेशी कुली जमैका टापू में दस बरस बराबर रहकर मिहनत का काम की सार्टिफिकेट पावे वा पाने के हकदार होवे तो वह कुली हर एक मरद के लिये राह खरच का आधा और औरत के लिये तीन हिस्सा का एक हिस्सा राह खरच दाखिल करने से टापू की तरफ से कलकत्ते तक पहुंचा दिया जायेगा। ऐसे कुली के लड़केबाले और जोरू और घराने अगर उसके साथ आवे तो बिना खरच आने सकेंगे।

पर जो कुली लाचार और बहुत गरीब हो वह लड़के बालोंको और जोरू और अपने घराने के लोगों को साथ लेकर बिना खरच लौट आने सकेगा। जो कुली एकबार टापू में से होकर हिन्दुस्तान आया या वह फिर लौट आने का खरच नहीं पावेगा। लगातार पांच बरस रहकर अगर कोई कुली मिहनत का काम करने के लिये सार्टिफिकेट पावे वा पाने का हकदार होवे तो अपने खरच से हिन्दुस्तान में लौट आने सकता है। आनेका वकत कम्बल और गरम कपड़ा मुफ़्त दिया जाता पर आने का समय नहीं।

अगर कोई कुली कोई समय टापू छोड़ दिया हो तो वह हिन्दुस्तान में लौट आने वकत टापू की तरफ़ से कुछ मदद नहीं पावेगा चाहे वह टापू में पूरा दश बरस रहा हो कि नहीं।

दूसरी शरतें—कालोनी या टापू में पहुंचने के पीछे हर एक परदेशी कुली को उसके मालीक की ओर से खाना जमैका गवर्मेंण्ट के ठहराए हुए हिसाब से पहिले तीन महीने भर हर हफता दो शिलिं छः पेन्स जो कि आजकाल एक रुपया चौदा आने के बराबर है देने से हर एक को जिसकी उमर बारह बरस या जियादे है मिलेगा।

हर एक लड़के या लड़की को जिसकी उमर एक और बारह बरस के भीतर है एक शिलिं तीन पैन्स जो कि आजकाल पंद्रह आनेका बराबर है देने से बड़ों का आधा खाना मिलेगा।

परदेशी कुलीयों को एकरार नामे के मुवाफ़िक रहने के लाएक घर बे किराए के मिलेगा जिसकी मरमत मालिक की ओर से अच्छी तरह होगी।

और जब परदेशी कुली लोग जो एकरार नामे के मुवाफ़िक हैं बीमार पड़ेंगे उन को हसपताल में जगह मिलेगी और डाक्टर दवाई बीमार योग्य खाना और मामूली खाना बिन दाम के मिलेगा।

Figure I.3: Districts that ever sent indentured immigrants to Fiji, Jamaica, Natal, or Suriname

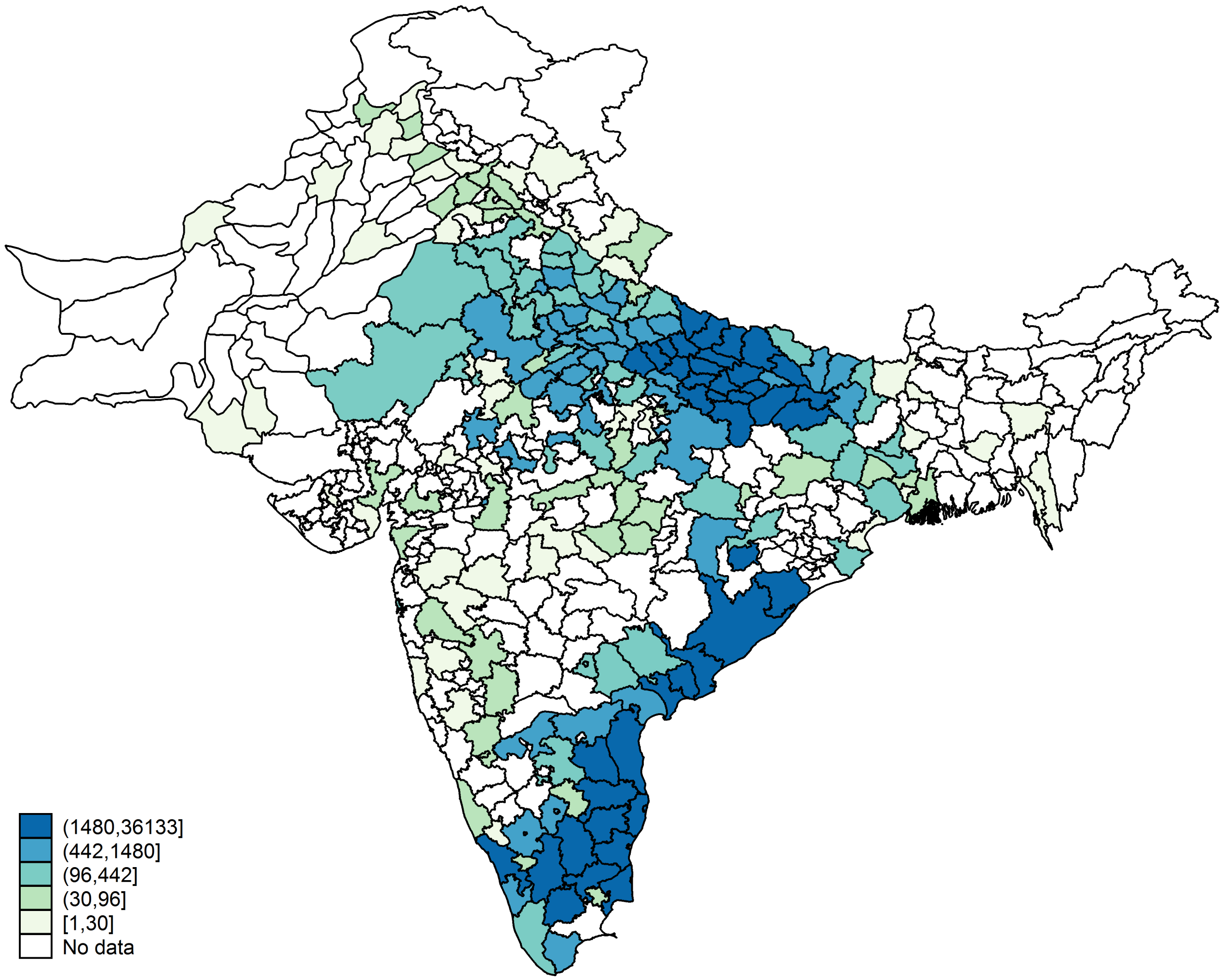


Figure I.4: Gorakhpur and Basti rice prices

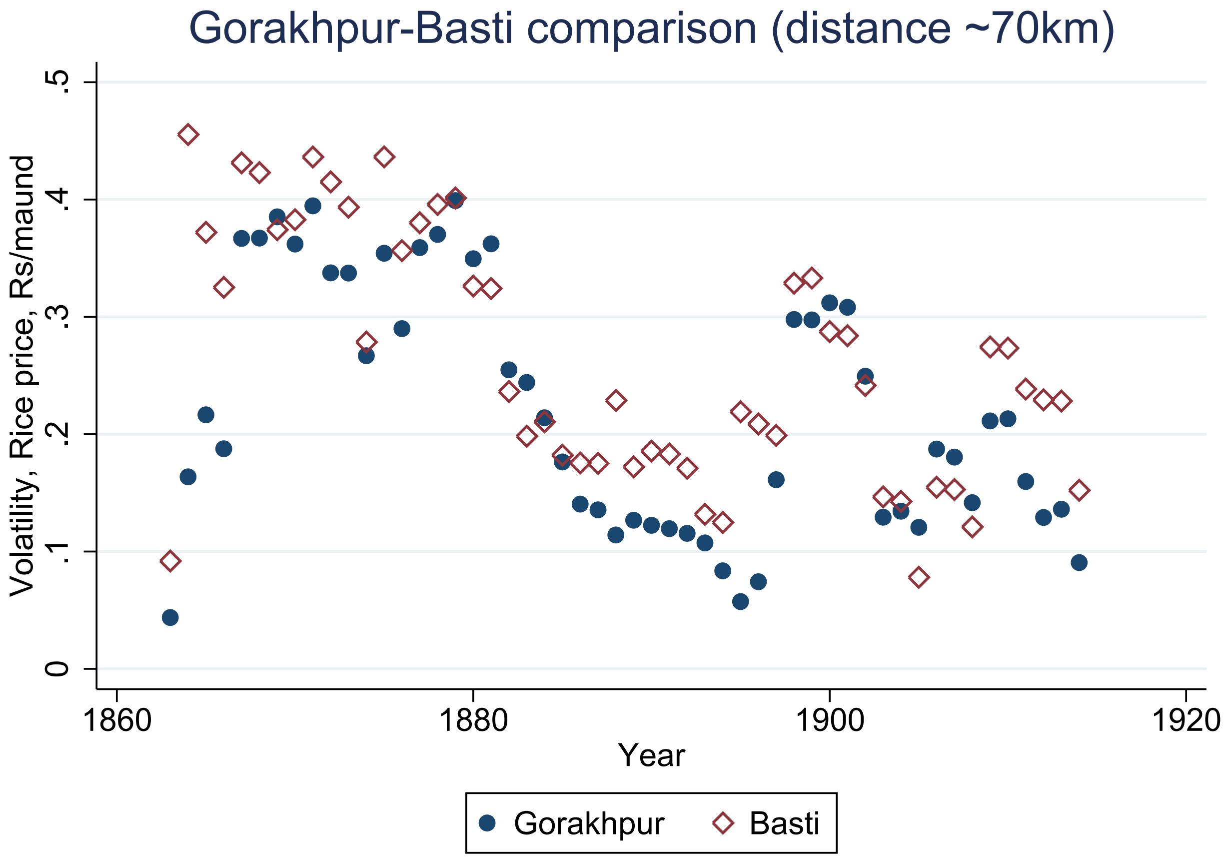


Figure I.5: Colonies vs India, monthly nominal wages (Rs)

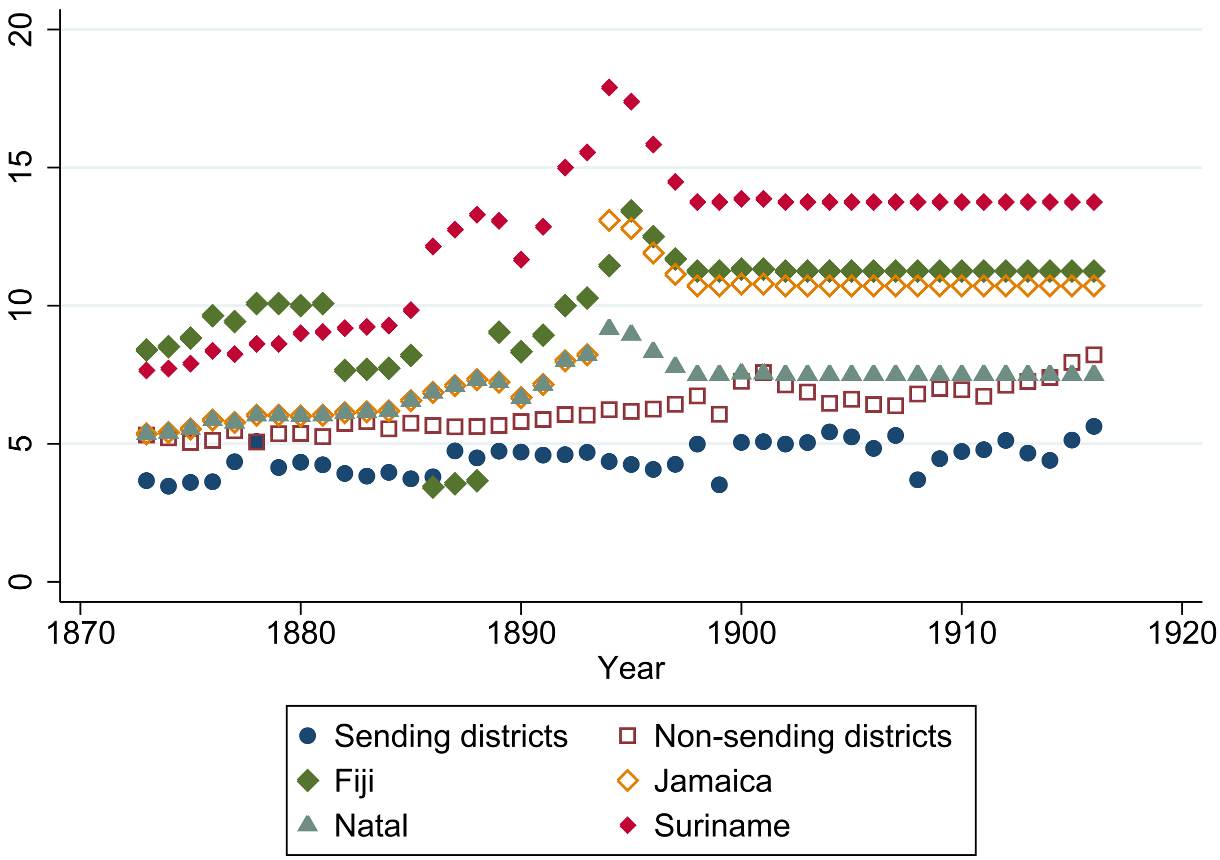


Table I.1: Estimated state-sponsored indentured immigration (India)

| Name of colony | Years of migration | Number of migrants |
|---|---|---|
| Mauritius | 1834-1900 | 453,063 |
| British Guiana | 1838-1916 | 238,909 |
| Trinidad | 1845-1916 | 143,939 |
| Jamaica | 1845-1915 | 36,412 |
| Grenada | 1856-1885 | 3,200 |
| St Lucia | 1858-1895 | 4,350 |
| Natal | 1860-1911 | 152,184 |
| St Kitts | 1860-1861 | 337 |
| St Vincent | 1860-1880 | 2,472 |
| Réunion | 1861-1883 | 26,507 |
| St Croix | 1863 - 1873 | 321 |
| Dutch Guiana | 1873-1916 | 34,304 |
| Fiji | 1879-1916 | 60,965 |
| East Africa | 1895-1916? | 32,000 |
| Seychelles | ?-1916 | 6,315 |
| *Total* | | 1,195,278 |

Source: Adapted from Lal (2004), Roopnarine (2010).

Table I.2: Dynamic Panel Check

| | (1) | (2) |
|---|---|---|
| Lag: log, Rice price, Rs/maund | 0.892*** | 0.698*** |
| | (0.00557) | (0.00718) |
| Constant | 0.169*** | 0.500*** |
| | (0.00777) | (0.0141) |
| Year FE | N | Y |
| Observations | 10595 | 10595 |

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.3: District summary statistics

| | Mean | SD | N |
|---|---|---|---|
| Sent any indentured immigrants to any colony | .54 | (.5) | 9452 |
| Number of indentured immigrants per district (conditional) | 45.55 | (146.55) | 5079 |
| Indentured immigrants per year | 5434.86 | (3369.54) | 42 |
| *All districts* | | | |
| Rice price (Rs/maund), all districts | 3.96 | (1.33) | 9330 |
| Volatility, Rice price, Rs/maund | .17 | (.1) | 9288 |
| Rice price (Rs/maund), rice districts | 3.78 | (1.17) | 4049 |
| Wheat price (Rs/maund), wheat districts | 2.93 | (.91) | 5279 |
| Wage Rs/month | 5.57 | (2.44) | 9276 |
| *Sending districts* | | | |
| Rice price (Rs/maund), all districts | 3.95 | (1.41) | 5031 |
| Rice price (Rs/maund), rice districts | 3.73 | (1.16) | 2045 |
| Wheat price (Rs/maund), wheat districts | 2.99 | (.87) | 2988 |
| Wage Rs/month | 5.07 | (2.21) | 4985 |
| *Non-sending districts* | | | |
| Rice price (Rs/maund), all districts | 3.98 | (1.23) | 4299 |
| Rice price (Rs/maund), rice districts | 3.84 | (1.18) | 2004 |
| Wheat price (Rs/maund), wheat districts | 2.85 | (.95) | 2291 |
| Wage Rs/month | 6.14 | (2.56) | 4291 |

Table I.4: Extensive margin with rice price

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| log, Min colony wage (Rs) | 0.191*** | 0.192*** | | |
| | (0.0243) | (0.0241) | | |
| log, Rice price, Rs/maund | 0.0457* | 0.0532** | -0.0475 | -0.0436 |
| | (0.0236) | (0.0244) | (0.0422) | (0.0425) |
| log, Wage Rs/month | -0.0324 | -0.0266 | -0.0530* | -0.0531 |
| | (0.0309) | (0.0332) | (0.0322) | (0.0335) |
| Volatility, Rice price, Rs/maund | | 0.205*** | | 0.141** |
| | | (0.0450) | | (0.0577) |
| Constant | 0.166** | 0.111 | 0.645*** | 0.625*** |
| | (0.0751) | (0.0799) | (0.0960) | (0.0978) |
| Year FEs | N | N | Y | Y |
| Observations | 9231 | 9179 | 9231 | 9179 |

Dependent variable: Sent any indentured immigrants to any colony.

SEs clustered by district in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.5: Intensive margin with rice price

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| log, Min colony wage (Rs) | 1.075*** | 1.084*** | | |
| | (0.118) | (0.118) | | |
| log, Rice price, Rs/maund | 0.233*** | 0.254*** | 0.308** | 0.314** |
| | (0.0769) | (0.0774) | (0.138) | (0.138) |
| log, Wage Rs/month | -0.202 | -0.188 | -0.150 | -0.148 |
| | (0.130) | (0.133) | (0.133) | (0.134) |
| Volatility, Rice price, Rs/maund | | 0.500** | | 0.273 |
| | | (0.199) | | (0.187) |
| Constant | 0.203 | 0.0554 | 1.352*** | 1.311*** |
| | (0.312) | (0.339) | (0.343) | (0.346) |
| Year FEs | N | N | Y | Y |
| Observations | 4970 | 4956 | 4970 | 4956 |

Dependent variable: Log, total indentured immigrants.

SEs clustered by district in parentheses. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.6: Counts with rice price

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Total indentured immigrants | | | | |
| log, Min colony wage (Rs) | 1.497*** | 1.620*** | | |
| | (0.181) | (0.213) | | |
| log, Rice price, Rs/maund | 0.617*** | 0.633*** | 0.893*** | 0.904*** |
| | (0.132) | (0.126) | (0.225) | (0.222) |
| log, Wage Rs/month | -0.232 | -0.191 | 0.106 | 0.105 |
| | (0.232) | (0.226) | (0.241) | (0.239) |
| Volatility, Rice price, Rs/maund | | 0.946* | | 0.892** |
| | | (0.544) | | (0.400) |
| Year FEs | N | N | Y | Y |
| Observations | 9101 | 9051 | 9101 | 9051 |

Dependent variable: Total indentured immigrants.

Bootstrapped SEs clustered by district in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.7: Extensive margin (SUR)

| | (1) | (2) | (3) |
|---|---|---|---|
| Fiji | | | |
| Log, fiji monthly wage (Rs) | 0.0859*** | 0.101*** | 0.0736*** |
| | -0.00421 | -0.00623 | -0.00514 |
| log, Rice price, Rs/maund | 0.0825*** | 0.140*** | 0.128*** |
| | -0.00428 | -0.00407 | -0.00484 |
| log, Wage Rs/month | 0.0509*** | 0.00273 | -0.00687 |
| | -0.00411 | -0.00363 | -0.00432 |
| Volatility, Rice price, Rs/maund | -0.0121 | 0.116*** | 0.135*** |
| | -0.00986 | -0.00983 | -0.0105 |
| Log, jamaica monthly wage (Rs) | | -0.176*** | |
| | | -0.012 | |
| Log, natal monthly wage (Rs) | | -0.0825** | |
| | | -0.0338 | |
| Log, suriname monthly wage (Rs) | | 0.267*** | |
| | | -0.0293 | |
| Jamaica | | | |
| Log, jamaica monthly wage (Rs) | 0.00931 | -0.00499 | 0.0681*** |
| | -0.00836 | -0.027 | -0.0089 |
| log, Rice price, Rs/maund | -0.0152 | 0.0198** | 0.0193* |
| | -0.00976 | -0.00998 | -0.00991 |
| log, Wage Rs/month | -0.0527*** | -0.0331*** | -0.0470*** |
| | -0.0083 | -0.00895 | -0.00853 |
| Volatility, Rice price, Rs/maund | -0.00804 | 0.0946*** | 0.0192 |
| | -0.0207 | -0.0219 | -0.0218 |
| Log, fiji monthly wage (Rs) | | -0.00482 | |
| | | -0.0132 | |
| Log, natal monthly wage (Rs) | | 0.237*** | |
| | | -0.0751 | |
| Log, suriname monthly wage (Rs) | | -0.0834 | |
| | | -0.0614 | |
| Natal | | | |
| Log, natal monthly wage (Rs) | 0.443*** | 0.446*** | 0.484*** |
| | -0.00697 | -0.0621 | -0.0078 |
| log, Rice price, Rs/maund | -0.289*** | -0.268*** | -0.278*** |
| | -0.00646 | -0.00754 | -0.00703 |
| log, Wage Rs/month | -0.194*** | -0.195*** | -0.188*** |
| | -0.00632 | -0.00683 | -0.00627 |
| Volatility, Rice price, Rs/maund | 0.471*** | 0.436*** | 0.468*** |
| | -0.0161 | -0.0178 | -0.0166 |
| Log, fiji monthly wage (Rs) | | 0.0882*** | |
| | | -0.0134 | |
| Log, jamaica monthly wage (Rs) | | -0.223*** | |
| | | -0.0215 | |
| Log, suriname monthly wage (Rs) | | 0.222*** | |
| | | -0.0511 | |
| Suriname | | | |
| Log, suriname monthly wage (Rs) | 0.00487 | -0.649*** | 0.0628*** |
| | -0.0072 | -0.0677 | -0.00807 |
| log, Rice price, Rs/maund | 0.0276*** | 0.00768 | 0.0720*** |
| | -0.00916 | -0.00985 | -0.00975 |
| log, Wage Rs/month | 0.0486*** | 0.000496 | -0.0401*** |
| | -0.00847 | -0.00882 | -0.00857 |
| Volatility, Rice price, Rs/maund | -0.117*** | -0.172*** | -0.0341 |
| | -0.0222 | -0.024 | -0.0233 |
| Log, fiji monthly wage (Rs) | | -0.0468*** | |
| | | -0.0146 | |
| Log, jamaica monthly wage (Rs) | | 0.0741*** | |
| | | -0.0287 | |
| Log, natal monthly wage (Rs) | | 1.018*** | |
| | | -0.0792 | |
| Own exclusion dummy | Y | Y | Y |
| Others' exclusion dummies | N | N | Y |
| Observations | 9179 | 9179 | 9179 |

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.8: North India summary statistics

| | Mean | SD | N |
|---|---|---|---|
| Age | 21.1 | (7.19) | 124075 |
| Height, male | 63.87 | (2.54) | 73276 |
| Height, female | 58.99 | (2.52) | 29819 |
| *Castes* | | | |
| Ahir | .1 | (.3) | 124361 |
| Brahman | .03 | (.17) | 124361 |
| Chamar | .14 | (.35) | 124361 |
| Kshatriya | .12 | (.32) | 124361 |
| Muslim | .12 | (.33) | 124361 |

Table I.9: Extensive margin (LPM) with rice price

| | (1) Brahman | (2) Kshatriya | (3) Ahir | (4) Chamar | (5) Muslim |
|---|---|---|---|---|---|
| log, Rice price, Rs/maund | 0.0460** | -0.0231 | 0.0384* | 0.0529** | 0.0469 |
| | (0.0212) | (0.0321) | (0.0227) | (0.0248) | (0.0324) |
| log, Wage Rs/month | -0.0861*** | 0.0117 | -0.0515*** | -0.0648*** | 0.00213 |
| | (0.0236) | (0.0327) | (0.0200) | (0.0232) | (0.0268) |
| Volatility, Rice price, Rs/maund | 0.00897 | 0.183** | 0.00290 | -0.00160 | 0.0532 |
| | (0.0444) | (0.0718) | (0.0738) | (0.0577) | (0.0641) |
| Constant | 0.112** | 0.235*** | 0.210*** | 0.181*** | 0.0998 |
| | (0.0540) | (0.0838) | (0.0629) | (0.0672) | (0.0775) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 7397 | 7397 | 7397 | 7397 | 7397 |

Dependent variable: any member of the caste listed. SEs clustered by district in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.10: Intensive margin with rice price

| | (1) Brahman | (2) Kshatriya | (3) Ahir | (4) Chamar | (5) Muslim |
|---|---|---|---|---|---|
| log, Rice price, Rs/maund | 0.581*** | -0.106 | 0.744*** | 0.700*** | 0.416*** |
| | (0.189) | (0.144) | (0.206) | (0.245) | (0.150) |
| log, Wage Rs/month | -0.177 | 0.105 | -0.601*** | -0.878*** | -0.458*** |
| | (0.186) | (0.178) | (0.227) | (0.298) | (0.140) |
| Volatility, Rice price, Rs/maund | -0.190 | 0.167 | -0.0598 | -0.0441 | 0.139 |
| | (0.386) | (0.201) | (0.345) | (0.620) | (0.252) |
| Constant | -0.337 | 0.494 | 0.518 | 0.941 | 0.580 |
| | (0.551) | (0.359) | (0.445) | (0.620) | (0.355) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 1067 | 2441 | 1673 | 1396 | 2395 |

Dependent variable: log of caste. SEs clustered by district in parentheses. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.11: Poisson regression with rice price

| | (1) Brahman | (2) Kshatriya | (3) Ahir | (4) Chamar | (5) Muslim |
|---|---|---|---|---|---|
| main | | | | | |
| log, Rice price, Rs/maund | 0.454 | -0.237 | 1.174*** | 1.639*** | 0.849*** |
| | (0.468) | (0.220) | (0.271) | (0.414) | (0.239) |
| log, Wage Rs/month | -0.200 | 0.263 | -0.453 | -0.729 | -0.221 |
| | (0.307) | (0.266) | (0.312) | (0.786) | (0.268) |
| Volatility, Rice price, Rs/maund | 0.0951 | 0.425 | -0.274 | -1.882 | 0.150 |
| | (0.563) | (0.305) | (0.665) | (1.483) | (0.442) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 4733 | 6206 | 4902 | 4161 | 6558 |

Dependent variable: count of each caste. Bootstrapped SEs clustered by district in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.12: One-by-one logit: caste shares

| | Male | | Female | | Children | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Brahman | | | | | | |
| log, Wage Rs/month | -0.0416*** | -0.00703 | -0.0604*** | -0.00947 | -0.0618*** | -0.0109 |
| | (0.0123) | (0.00784) | (0.0157) | (0.0126) | (0.0183) | (0.0163) |
| log, Colony wage Rs | -0.0318*** | 0.0250*** | -0.0421*** | 0.0295* | -0.0200* | 0.0396 |
| | (0.00663) | (0.00848) | (0.00895) | (0.0157) | (0.0120) | (0.0309) |
| log, Rice price, Rs/maund | -0.00202 | -0.00278 | -0.0163** | -0.0275** | -0.0126* | -0.0299** |
| | (0.00241) | (0.00807) | (0.00818) | (0.0107) | (0.00639) | (0.0149) |
| Volatility, Rice price, Rs/maund | -0.00304 | -0.00108 | -0.0808*** | -0.0718** | -0.0279 | -0.00833 |
| | (0.0107) | (0.0124) | (0.0215) | (0.0296) | (0.0237) | (0.0296) |
| Kshatriya | | | | | | |
| log, Wage Rs/month | 0.0264 | 0.0140 | 0.0552** | 0.0317 | 0.0715* | 0.0698 |
| | (0.0179) | (0.0168) | (0.0267) | (0.0284) | (0.0387) | (0.0462) |
| log, Colony wage Rs | 0.0719*** | 0.0269 | 0.0830*** | 0.0632* | 0.0710*** | 0.0499* |
| | (0.00837) | (0.0193) | (0.0138) | (0.0354) | (0.0158) | (0.0288) |
| log, Rice price, Rs/maund | -0.0522*** | -0.0754*** | -0.0475*** | -0.0837*** | -0.0104 | -0.0870*** |
| | (0.00825) | (0.0158) | (0.0122) | (0.0315) | (0.0154) | (0.0322) |
| Volatility, Rice price, Rs/maund | 0.00623 | 0.0290 | 0.0270 | 0.0221 | 0.0540 | 0.0722 |
| | (0.0219) | (0.0215) | (0.0344) | (0.0379) | (0.0726) | (0.0696) |
| Ahir | | | | | | |
| log, Wage Rs/month | 0.00651 | -0.0135 | 0.0106 | -0.0107 | -0.0157 | -0.0438 |
| | (0.0146) | (0.0103) | (0.0246) | (0.0219) | (0.0271) | (0.0272) |
| log, Colony wage Rs | 0.0534*** | -0.00811 | 0.0691*** | 0.0362 | 0.0522** | -0.0439 |
| | (0.0122) | (0.0170) | (0.0163) | (0.0272) | (0.0201) | (0.0317) |
| log, Rice price, Rs/maund | 0.0206** | 0.0161* | 0.00559 | 0.00390 | 0.0122 | -0.0339 |
| | (0.00938) | (0.00825) | (0.0110) | (0.0123) | (0.0126) | (0.0221) |
| Volatility, Rice price, Rs/maund | -0.0341* | 0.0101 | 0.0159 | -0.00651 | -0.0271 | -0.00789 |
| | (0.0204) | (0.0250) | (0.0256) | (0.0298) | (0.0321) | (0.0473) |
| Chamar | | | | | | |
| log, Wage Rs/month | -0.0225 | -0.0213 | -0.0495 | -0.0474 | -0.0562 | -0.0641 |
| | (0.0184) | (0.0172) | (0.0449) | (0.0354) | (0.0801) | (0.0536) |
| log, Colony wage Rs | 0.0424*** | 0.0157 | 0.0329** | -0.0281 | 0.0549** | -0.0727 |
| | (0.0116) | (0.0200) | (0.0154) | (0.0269) | (0.0241) | (0.0559) |
| log, Rice price, Rs/maund | 0.0416*** | 0.0325** | 0.0703*** | 0.0929*** | 0.0513** | 0.161*** |
| | (0.00549) | (0.0136) | (0.0125) | (0.0277) | (0.0244) | (0.0498) |
| Volatility, Rice price, Rs/maund | -0.106*** | -0.00749 | -0.237*** | -0.0752 | -0.312*** | -0.0101 |
| | (0.0349) | (0.0267) | (0.0633) | (0.0670) | (0.0990) | (0.126) |
| Muslim | | | | | | |
| log, Wage Rs/month | -0.0222 | 0.0243* | -0.0481* | -0.0155 | -0.0281 | 0.000856 |
| | (0.0154) | (0.0140) | (0.0266) | (0.0288) | (0.0329) | (0.0396) |
| log, Colony wage Rs | 0.0210** | 0.135*** | 0.00727 | 0.105*** | 0.0360** | 0.168*** |
| | (0.00956) | (0.0233) | (0.0133) | (0.0303) | (0.0180) | (0.0424) |
| log, Rice price, Rs/maund | -0.0104 | 0.00518 | -0.0266* | -0.00559 | -0.0164 | 0.0272 |
| | (0.00861) | (0.0130) | (0.0140) | (0.0231) | (0.0128) | (0.0348) |
| Volatility, Rice price, Rs/maund | 0.0914*** | 0.0146 | 0.0670** | -0.0840 | 0.0977* | -0.0646 |
| | (0.0197) | (0.0303) | (0.0320) | (0.0560) | (0.0558) | (0.0704) |
| District FEs | Y | Y | Y | Y | Y | Y |
| Colony FEs | Y | Y | Y | Y | Y | Y |
| Yr FEs | N | Y | N | Y | N | Y |
| Observations | 77030 | 77030 | 31562 | 31562 | 12562 | 12562 |

SEs clustered by district in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.13: Madras-Calcutta height comparison

| | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Madras | 0.253$^{*}$ | 0.199 | 0.0733 | -0.00952 | 0.672$^{***}$ | 0.621$^{***}$ | 0.706$^{***}$ | 0.622$^{***}$ |
| | (0.145) | (0.145) | (0.137) | (0.134) | (0.123) | (0.113) | (0.137) | (0.123) |
| Year FEs | No | Yes | No | Yes | No | Yes | No | Yes |
| Age controls | No | No | Yes | Yes | No | No | Yes | Yes |
| Observations | 142356 | 142356 | 142314 | 142314 | 56087 | 56087 | 56052 | 56052 |

Dependent variable: height (cms). SEs clustered by district in parentheses. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.14: Height regressions, Male

| | Both ports | | Calcutta | | Madras | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log, Wage Rs/month | 0.00339** | 0.00341** | 0.00355** | 0.00360** | 0.00268* | 0.00265* |
| | (0.0014) | (0.0014) | (0.0018) | (0.0018) | (0.0015) | (0.0015) |
| log, Colony wage | -0.00941*** | -0.00941*** | -0.00707*** | -0.00707*** | 0.0863*** | 0.0868*** |
| | (0.0019) | (0.0019) | (0.0027) | (0.0027) | (0.0055) | (0.0056) |
| log, Rice price | -0.00368*** | -0.00372*** | -0.00392*** | -0.00398*** | -0.00172 | -0.00173 |
| | (0.00097) | (0.00098) | (0.0011) | (0.0011) | (0.0036) | (0.0036) |
| Madras | -0.00372* | -0.00375* | | | | |
| | (0.0020) | (0.0021) | | | | |
| Age | 0.00247*** | 0.00247*** | 0.00655*** | 0.00655*** | 0.00261*** | 0.00261*** |
| | (0.00019) | (0.00019) | (0.00056) | (0.00056) | (0.00016) | (0.00016) |
| $Age^2$ | -0.0000104*** | -0.0000104*** | -0.0000945*** | -0.0000946*** | -0.0000102*** | -0.0000102*** |
| | (0.0000028) | (0.0000028) | (0.000010) | (0.000011) | (0.0000024) | (0.0000024) |
| Volatility, Rice price | | 0.000443 | | -0.000170 | | -0.00118 |
| | | (0.0016) | | (0.0019) | | (0.0028) |
| Constant | 4.131*** | 4.131*** | 4.073*** | 4.073*** | 3.896*** | 3.895*** |
| | (0.0068) | (0.0069) | (0.0085) | (0.0085) | (0.017) | (0.017) |
| $N$ | 137765 | 137724 | 72093 | 72054 | 65674 | 65672 |

Dependent variable: log, height (ins). SEs clustered by district in parentheses. Rice prices for Rs/maund. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.15: Height regressions, Female

| | Both ports | | Calcutta | | Madras | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log, Wage Rs/month | -0.00392** | -0.00401** | -0.00231 | -0.00266 | -0.00501** | -0.00484** |
| | (0.0017) | (0.0017) | (0.0025) | (0.0024) | (0.0021) | (0.0021) |
| log, Colony wage | -0.0114*** | -0.0112*** | -0.0109** | -0.0108** | -0.0108*** | -0.0108*** |
| | (0.0032) | (0.0032) | (0.0042) | (0.0042) | (0.0024) | (0.0024) |
| log, Rice price | -0.00577*** | -0.00555*** | -0.00246 | -0.00267 | -0.0109*** | -0.0108*** |
| | (0.0020) | (0.0020) | (0.0026) | (0.0027) | (0.0038) | (0.0040) |
| Madras | 0.00505* | 0.00437 | | | | |
| | (0.0027) | (0.0026) | | | | |
| Age | 0.00316*** | 0.00317*** | 0.00221*** | 0.00220*** | 0.00601*** | 0.00601*** |
| | (0.00068) | (0.00068) | (0.00021) | (0.00021) | (0.0014) | (0.0014) |
| $Age^2$ | -0.0000271** | -0.0000271** | -0.0000139*** | -0.0000139*** | -0.0000774*** | -0.0000774*** |
| | (0.000012) | (0.000012) | (0.0000033) | (0.0000033) | (0.000027) | (0.000027) |
| Volatility, Rice price | | 0.0119*** | | 0.00809 | | 0.00747 |
| | | (0.0037) | | (0.0052) | | (0.0049) |
| Constant | 4.065*** | 4.063*** | 4.064*** | 4.063*** | 4.050*** | 4.048*** |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.020) | (0.020) |
| $N$ | 55407 | 55399 | 29316 | 29309 | 26091 | 26090 |

Dependent variable: log, height (ins). SEs clustered by district in parentheses. Rice prices for Rs/maund. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.16: Return migration and mortality summary statistics

| | Mean | SD | N |
|---|---|---|---|
| *Repatriated* | | | |
| *Calcutta* | | | |
| All | .35 | (.48) | 26607 |
| Male adult | .38 | (.49) | 16990 |
| Female adult | .29 | (.46) | 6976 |
| Child | .29 | (.45) | 2550 |
| Duration (days) | 3325.45 | (1841.46) | 9259 |
| *Castes* | | | |
| Ahir | .32 | (.47) | 2423 |
| Brahman | .49 | (.5) | 780 |
| Chamar | .28 | (.45) | 3118 |
| Kshatriya | .37 | (.48) | 3947 |
| Muslim | .4 | (.49) | 3711 |
| *Madras* | | | |
| All | .24 | (.42) | 598 |
| Male adult | .22 | (.41) | 405 |
| Female adult | .24 | (.43) | 144 |
| Child | .37 | (.49) | 49 |
| Duration (days) | 2000.65 | (1375.31) | 141 |
| *Mortality* | | | |
| *Calcutta* | | | |
| All | .18 | (.38) | 26607 |
| Male adult | .16 | (.36) | 16990 |
| Female adult | .2 | (.4) | 6976 |
| Child | .24 | (.43) | 2550 |
| Duration (days) | 2113.73 | (2447.83) | 4694 |
| *Castes* | | | |
| Ahir | .19 | (.39) | 2423 |
| Brahman | .22 | (.42) | 780 |
| Chamar | .18 | (.38) | 3118 |
| Kshatriya | .18 | (.38) | 3947 |
| Muslim | .15 | (.36) | 3711 |
| *Madras* | | | |
| All | .1 | (.31) | 598 |
| Male adult | .09 | (.29) | 405 |
| Female adult | .1 | (.3) | 144 |
| Child | .2 | (.41) | 49 |
| Duration (days) | 951.34 | (1010.45) | 62 |

Table I.17: Survival in Fiji, failure=repatriated

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Female | -0.247*** | -0.246** | -0.260*** | -0.403*** | -0.155** |
| | (-6.60) | (-3.20) | (-5.98) | (-5.76) | (-2.74) |
| Age | 0.0190*** | 0.0520*** | 0.0227 | -0.022 | 0.0730* |
| | (4.52) | (6.34) | (1.04) | (-0.73) | (2.15) |
| Age$^2$ | 0.0000195 | -0.000159** | -0.000158 | 0.000743 | -0.00123 |
| | (0.50) | (-3.14) | (-0.38) | (1.34) | (-1.84) |
| Height (in) | -0.0349 | -0.165 | -0.0765 | -0.0953* | 0.226 |
| | (-0.91) | (-1.70) | (-1.54) | (-2.10) | (1.40) |
| Height$^2$ | 0.000315 | 0.000973 | 0.000758* | 0.000914** | -0.00167 |
| | (1.07) | (1.29) | (1.99) | (2.82) | (-1.30) |
| log, Rice price, Rs/maund | 0.133 | -0.41 | 0.243* | 0.135 | 0.328* |
| | (1.31) | (-1.89) | (2.02) | (0.64) | (2.17) |
| Volatility, Rice price, Rs/maund | -0.0337 | -0.856 | 0.238 | 0.0718 | 0.284 |
| | (-0.13) | (-1.70) | (0.79) | (0.17) | (0.66) |
| log, Wage Rs/month t | -0.0421 | 0.547 | -0.239 | -0.212 | -0.155 |
| | (-0.33) | (1.86) | (-1.62) | (-0.90) | (-0.81) |
| Ahir | 0.147** | 0.0368 | 0.183** | 0.315** | 0.0995 |
| | (2.74) | (0.35) | (2.92) | (3.14) | (1.23) |
| Brahman | 0.479*** | 0.491** | 0.524*** | 0.674*** | 0.368** |
| | (6.72) | (3.14) | (6.48) | (5.73) | (3.19) |
| Muslim | 0.323*** | 0.0727 | 0.385*** | 0.295** | 0.464*** |
| | (6.68) | (0.74) | (6.85) | (3.16) | (6.58) |
| Kshatriya | 0.389*** | 0.216* | 0.428*** | 0.524*** | 0.385*** |
| | (7.90) | (2.24) | (7.43) | (5.61) | (5.23) |
| Observations | 22096 | 22096 | 18136 | 18136 | 14966 |
| Time | All | First 5 | After 5 | 5-10 | After 10 |

Standard errors in parentheses. Other caste binaries and district FEs not reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.18: Survival in Fiji, failure=death

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Female | 0.177*** | 0.286*** | 0.0141 | 0.046 | -0.0161 |
| | (3.47) | (4.41) | (0.17) | (0.38) | (-0.14) |
| Age | 0.0656** | 0.0717* | 0.0612 | 0.0962 | 0.052 |
| | (2.91) | (2.20) | (1.84) | (1.59) | (1.29) |
| $\text{Age}^2$ | -0.000413 | -0.000587 | -0.00022 | -0.000873 | 0.0000291 |
| | (-1.00) | (-0.97) | (-0.37) | (-0.77) | (0.04) |
| Height (in) | -0.0365 | -0.0217 | -0.0682 | -0.222 | 0.199 |
| | (-0.52) | (-0.27) | (-0.33) | (-1.49) | (0.61) |
| $\text{Height}^2$ | 0.000214 | 0.000133 | 0.000411 | 0.00148 | -0.0016 |
| | (0.39) | (0.22) | (0.25) | (1.28) | (-0.61) |
| log, Rice price, Rs/maund | 0.142 | 0.171 | 0.173 | -0.0824 | 0.465 |
| | (0.94) | (0.89) | (0.68) | (-0.21) | (1.28) |
| Volatility, Rice price, Rs/maund | 0.00695 | 0.527 | -1.115 | -1.957* | -0.0993 |
| | (0.02) | (1.23) | (-1.63) | (-2.03) | (-0.09) |
| log, Wage Rs/month t | -0.215 | -0.0446 | -0.47 | -0.885 | -0.27 |
| | (-1.14) | (-0.18) | (-1.54) | (-1.86) | (-0.66) |
| Ahir | 0.0943 | 0.153 | -0.00283 | 0.0934 | -0.0879 |
| | (1.32) | (1.67) | (-0.02) | (0.55) | (-0.57) |
| Brahman | 0.013 | 0.198 | -0.422* | -0.312 | -0.474 |
| | (0.13) | (1.61) | (-2.18) | (-1.01) | (-1.90) |
| Muslim | -0.212** | -0.274** | -0.149 | -0.0473 | -0.216 |
| | (-3.03) | (-2.96) | (-1.38) | (-0.29) | (-1.45) |
| Kshatriya | 0.0658 | 0.121 | -0.00381 | 0.094 | -0.0995 |
| | (0.98) | (1.40) | (-0.04) | (0.59) | (-0.66) |
| Observations | 22096 | 22096 | 18136 | 18136 | 14966 |
| Time | All | First 5 | After 5 | 5-10 | After 10 |

Standard errors in parentheses. Chamar is the omitted caste. Other caste binaries and district FEs not reported.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.19: Survival in Fiji, failure=repatriated

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Female | -0.212*** | -0.195*** | -0.201*** | -0.191*** | -0.197*** |
| | (0.0315) | (0.0324) | (0.0326) | (0.0330) | (0.0332) |
| Departure year | -0.0280*** | -0.0280*** | -0.0255*** | -0.0361*** | -0.0349*** |
| | (0.00283) | (0.00304) | (0.00308) | (0.00549) | (0.00551) |
| Departure age | 0.0117** | 0.0136** | 0.0127** | 0.0157*** | 0.0146*** |
| | (0.00541) | (0.00542) | (0.00545) | (0.00548) | (0.00551) |
| Departure age$^2$ | 0.0000404 | 0.0000337 | 0.0000432 | 0.0000111 | 0.0000210 |
| | (0.0000520) | (0.0000502) | (0.0000506) | (0.0000499) | (0.0000504) |
| Ahir | | | 0.0620 | | 0.0685 |
| | | | (0.0675) | | (0.0677) |
| Brahman | | | 0.419*** | | 0.399*** |
| | | | (0.0777) | | (0.0799) |
| Muslim | | | 0.183*** | | 0.183*** |
| | | | (0.0603) | | (0.0607) |
| Kshatriya | | | 0.294*** | | 0.281*** |
| | | | (0.0609) | | (0.0621) |
| log, Fiji wage | | | | 1.150*** | 1.150*** |
| | | | | (0.0799) | (0.0799) |
| log, India wage | | | | 0.755*** | 0.743*** |
| | | | | (0.141) | (0.141) |
| log, Rice price | | | | -0.791*** | -0.790*** |
| | | | | (0.104) | (0.104) |
| Volatility, Rice price | | | | -2.404*** | -2.395*** |
| | | | | (0.239) | (0.239) |
| log, Rice price (depart yr) | | | | -0.330*** | -0.302*** |
| | | | | (0.107) | (0.107) |
| log, Fiji wage (depart yr) | | | | 0.143*** | 0.157*** |
| | | | | (0.0523) | (0.0524) |
| log, India wage (depart yr) | | | | 0.289** | 0.318** |
| | | | | (0.139) | (0.140) |
| Volatility, Rice price (depart yr) | | | | -0.528** | -0.540** |
| | | | | (0.238) | (0.238) |
| Observations | 153901 | 152171 | 152171 | 147023 | 147023 |
| District FEs | N | Y | Y | Y | Y |

Standard errors in parentheses. Other caste binaries and district FEs not reported.

Rice prices measured in Rs/maund. Wages in Rs/month. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Figure I.A1: Distribution of log rice price (Rs)

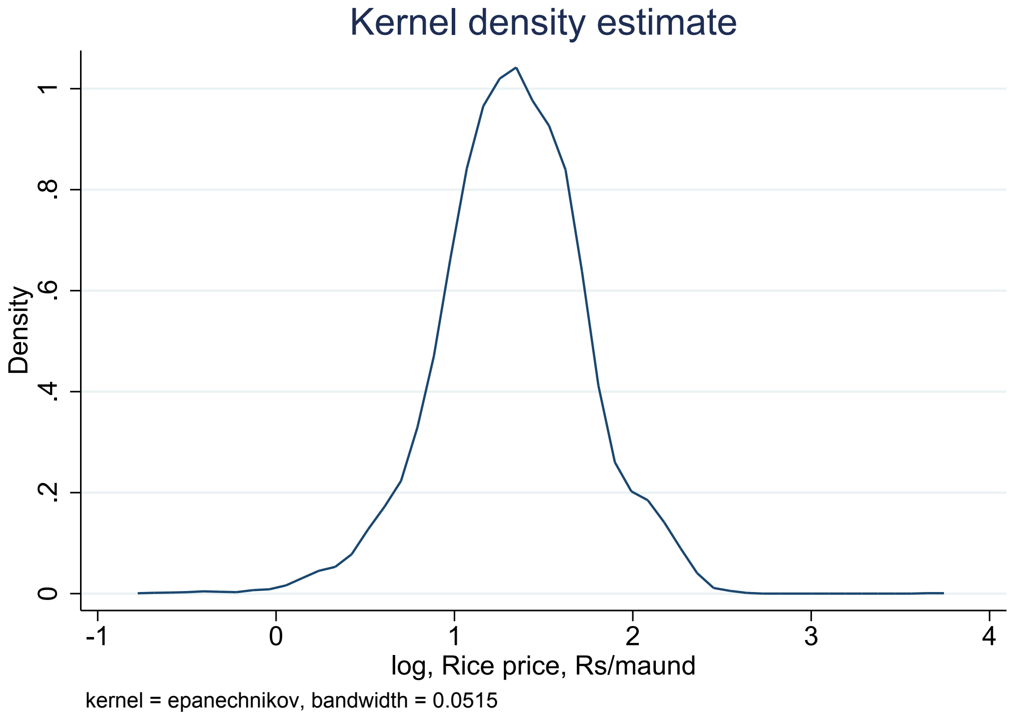


Figure I.A2: Distribution of log wage (Rs)

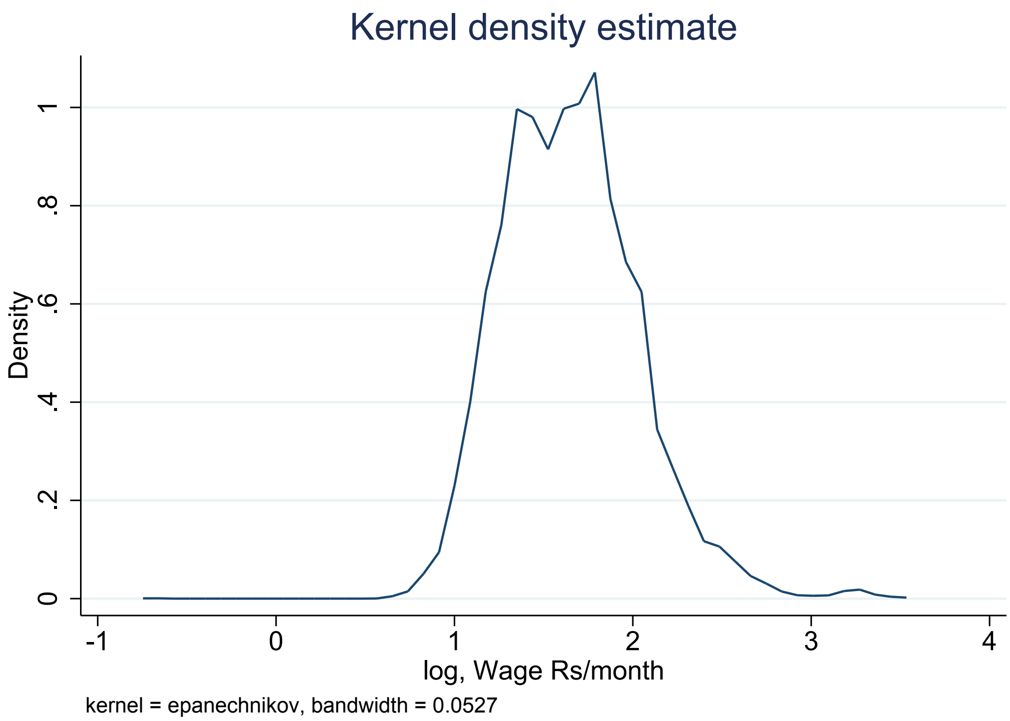


Figure I.A3: Height distributions (in), males

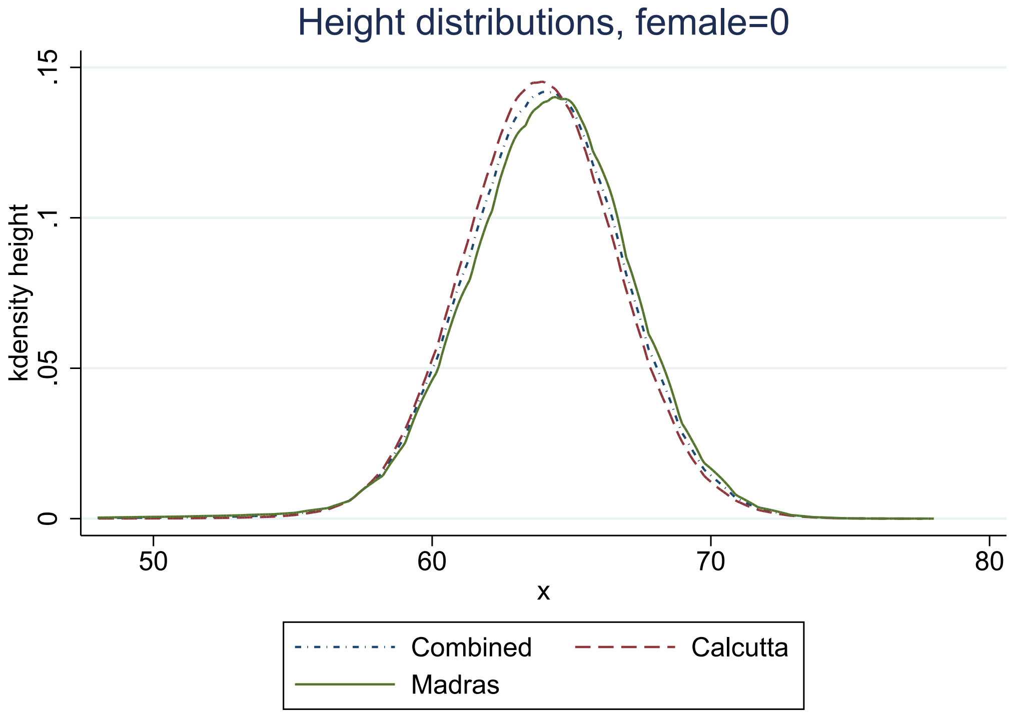


Figure I.A4: Height distributions (in), females

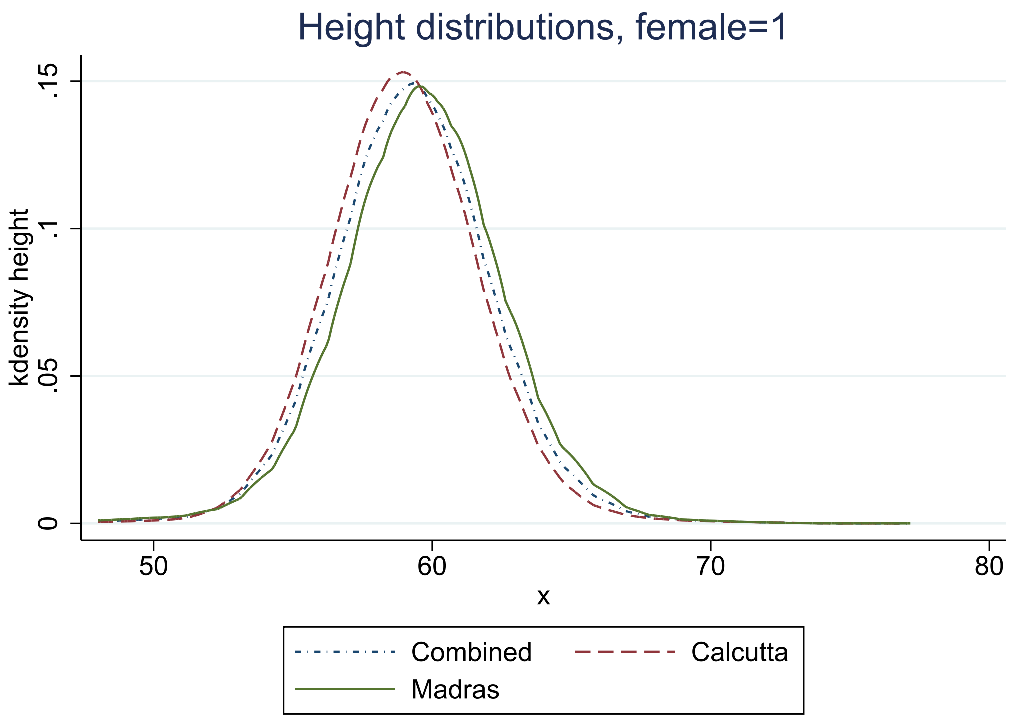


Table I.A1: Check against annual reports (UP/Bihar)

| | (1) | (2) | (3) |
|---|---|---|---|
| Sent any indentured immigrants to any colony | 0.985*** | | |
| | (0.00973) | | |
| Total indentured immigrants | | 4.906*** | 4.908*** |
| | | (0.0772) | (0.0792) |
| Sample | All | All | Sending |
| $R^2$ | 0.906 | 0.792 | 0.792 |
| Observations | 1064 | 1064 | 1009 |

Constant dropped from the regression. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A2: Extensive margin (probit) with rice price (RE probit)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Sent any indentured immigrants to any colony | | | | |
| Min colony wage (Rs) | 0.974*** | 0.982*** | | |
| | (0.108) | (0.113) | | |
| log, Rice price, Rs/maund | 0.182** | 0.217** | -0.248 | -0.237 |
| | (0.0902) | (0.0935) | (0.163) | (0.185) |
| log, Wage Rs/month | -0.277** | -0.260* | -0.537*** | -0.548*** |
| | (0.117) | (0.139) | (0.146) | (0.158) |
| Volatility, Rice price, Rs/maund | | 0.928*** | | 0.726** |
| | | (0.199) | | (0.364) |
| Year FEs | N | N | Y | Y |
| Observations | 9187 | 9135 | 9187 | 9135 |

Dependent variable: any indentured immigrant. Bootstrapped SEs clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A3: Negative binomial with rice price

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Total indentured immigrants | | | | |
| Min colony wage (Rs) | 1.088*** | 1.112*** | | |
| | (0.111) | (0.121) | | |
| log, Rice price, Rs/maund | 0.333*** | 0.360*** | 0.410* | 0.439** |
| | (0.103) | (0.0986) | (0.222) | (0.217) |
| log, Wage Rs/month | -0.765*** | -0.747*** | -0.660*** | -0.663*** |
| | (0.159) | (0.161) | (0.142) | (0.158) |
| Volatility, Rice price, Rs/maund | | 0.577*** | | 0.312 |
| | | (0.192) | | (0.213) |
| Constant | -1.839*** | -2.037*** | -0.320 | -0.394 |
| | (0.258) | (0.308) | (0.487) | (0.472) |
| Year FEs | N | N | Y | Y |
| Observations | 9060 | 9010 | 9060 | 9010 |

Dependent variable: Total indentured immigrants. Bootstrapped SEs clustered by state in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A4: Check using data from annual reports (UP/Bihar)

| | Extensive | | Intensive | | Poisson | | Negative binomial | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| main | | | | | | | | |
| Min colony wage (Rs) | 0.173*** | | 0.803*** | | 0.790*** | | 0.606*** | |
| | (0.0226) | | (0.107) | | (0.199) | | (0.0929) | |
| log, Rice price | 0.0522** | -0.0428 | 0.218*** | 0.259* | 0.334*** | 0.792*** | 0.357*** | 0.555*** |
| | (0.0243) | (0.0436) | (0.0843) | (0.150) | (0.0906) | (0.213) | (0.0912) | (0.177) |
| log, Wage Rs/month | -0.0206 | -0.0551* | -0.0839 | -0.0752 | 0.0887 | 0.0497 | -0.424*** | -0.287** |
| | (0.0320) | (0.0329) | (0.146) | (0.137) | (0.205) | (0.242) | (0.134) | (0.129) |
| Volatility, Rice price | 0.190*** | 0.139** | 0.358 | 0.291 | -0.612 | 0.960* | 0.197 | 0.345 |
| | (0.0458) | (0.0579) | (0.223) | (0.185) | (0.607) | (0.500) | (0.255) | (0.280) |
| Constant | 0.142* | 0.626*** | 0.469 | 1.265*** | | | -1.735*** | -1.428*** |
| | (0.0782) | (0.0991) | (0.360) | (0.363) | | | (0.297) | (0.389) |
| Year FE | N | Y | N | Y | | | | |
| Observations | 9135 | 9135 | 5008 | 5008 | 9010 | 9010 | 9010 | 9010 |

Data for Bihar and UP replaced by annual reports. Rice price in Rs/maund. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A5: Robustness by excluding peripheral areas

| | Extensive | | Intensive | | Poisson | | Negative binomial | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| main | | | | | | | | |
| Min colony wage (Rs) | 0.199*** | | 1.102*** | | 1.662*** | | 1.107*** | |
| | (0.0249) | | (0.117) | | (0.225) | | (0.116) | |
| log, Rice price | 0.0441* | -0.0533 | 0.235*** | 0.302** | 0.589*** | 0.877*** | 0.356*** | 0.439** |
| | (0.0244) | (0.0446) | (0.0752) | (0.133) | (0.130) | (0.212) | (0.0964) | (0.210) |
| log, Wage Rs/month | -0.0223 | -0.0503 | -0.163 | -0.130 | -0.173 | 0.110 | -0.747*** | -0.664*** |
| | (0.0317) | (0.0346) | (0.141) | (0.135) | (0.244) | (0.242) | (0.161) | (0.148) |
| Volatility, Rice price | 0.217*** | 0.140** | 0.506** | 0.254 | 0.976* | 0.835** | 0.583*** | 0.318 |
| | (0.0469) | (0.0564) | (0.203) | (0.203) | (0.530) | (0.389) | (0.192) | (0.201) |
| Constant | 0.116 | 0.646*** | 0.0171 | 1.305*** | | | -2.023*** | -0.396 |
| | (0.0795) | (0.100) | (0.344) | (0.339) | | | (0.302) | (0.448) |
| Year FE | N | Y | N | Y | N | Y | N | Y |
| Observations | 8828 | 8828 | 4908 | 4908 | 8732 | 8732 | 8732 | 8732 |

Rice price measured in Rs/maund. Excludes Assam, Berar, Coorg, Baluchistan, and Sind. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A6: Robustness with wage differential

| | Extensive | | Intensive | | Poisson | | Negative binomial | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| main | | | | | | | | |
| log, Rice price | 0.0971*** | -0.0433 | 0.480*** | 0.306** | 0.856*** | 0.847*** | 0.419*** | 0.430** |
| | (0.0244) | (0.0437) | (0.0919) | (0.140) | (0.125) | (0.210) | (0.0876) | (0.209) |
| Log wage differential | 0.0115 | -0.00830 | 0.126*** | -0.0124 | 0.174*** | -0.132** | 0.253*** | 0.164*** |
| | (0.00748) | (0.00835) | (0.0335) | (0.0361) | (0.0669) | (0.0669) | (0.0471) | (0.0561) |
| Volatility, Rice price | 0.185*** | 0.118** | 0.391* | 0.232 | 0.505 | 0.862** | 0.573*** | 0.251 |
| | (0.0445) | (0.0564) | (0.220) | (0.181) | (0.515) | (0.381) | (0.169) | (0.217) |
| Constant | 0.389*** | 0.538*** | 1.535*** | 1.073*** | | | -1.262*** | -1.555*** |
| | (0.0443) | (0.0823) | (0.158) | (0.284) | | | (0.161) | (0.381) |
| Year FE | N | Y | N | Y | N | Y | N | Y |
| Observations | 9094 | 9094 | 4929 | 4929 | 8972 | 8972 | 8972 | 8972 |

Includes a dummy for negative differential. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A7: Extensive margin (LPM) with rice price, by sex/age

| | (1) Brahman any | (2) Kshatriya any | (3) Ahir any | (4) Chamar any | (5) Muslim any |
|---|---|---|---|---|---|
| *Male adults* | | | | | |
| log, Rice price, Rs/maund | 0.0383* | -0.0431 | 0.0422* | 0.0460* | 0.0317 |
| | (0.0205) | (0.0318) | (0.0232) | (0.0244) | (0.0335) |
| log, Wage Rs/month | -0.0703*** | 0.0189 | -0.0477** | -0.0543** | -0.0166 |
| | (0.0230) | (0.0325) | (0.0196) | (0.0225) | (0.0269) |
| Volatility, Rice price, Rs/maund | 0.00555 | 0.183** | 0.00988 | 0.00573 | 0.0713 |
| | (0.0425) | (0.0739) | (0.0735) | (0.0598) | (0.0744) |
| Constant | 0.0792 | 0.232*** | 0.197*** | 0.161** | 0.111 |
| | (0.0538) | (0.0842) | (0.0634) | (0.0664) | (0.0797) |
| *Female adults* | | | | | |
| log, Rice price, Rs/maund | 0.0350** | -0.0000312 | 0.0502*** | 0.0654*** | 0.0631** |
| | (0.0136) | (0.0262) | (0.0175) | (0.0249) | (0.0249) |
| log, Wage Rs/month | -0.0758*** | -0.0574** | -0.0418** | -0.0705*** | -0.0712*** |
| | (0.0193) | (0.0273) | (0.0196) | (0.0225) | (0.0268) |
| Volatility, Rice price, Rs/maund | -0.0526* | -0.0150 | -0.0323 | -0.0508 | -0.0158 |
| | (0.0300) | (0.0541) | (0.0308) | (0.0414) | (0.0605) |
| Constant | 0.112*** | 0.211*** | 0.0658 | 0.116* | 0.146** |
| | (0.0374) | (0.0577) | (0.0456) | (0.0604) | (0.0625) |
| *Children* | | | | | |
| log, Rice price, Rs/maund | 0.0302*** | 0.00797 | 0.0399** | 0.0411** | 0.0713*** |
| | (0.00963) | (0.0157) | (0.0157) | (0.0186) | (0.0179) |
| log, Wage Rs/month | -0.0341*** | -0.0137 | -0.0277* | -0.0442*** | -0.0455*** |
| | (0.0102) | (0.0193) | (0.0148) | (0.0161) | (0.0148) |
| Volatility, Rice price, Rs/maund | -0.0321 | -0.000314 | -0.0279 | -0.0333 | -0.0394 |
| | (0.0204) | (0.0340) | (0.0264) | (0.0238) | (0.0299) |
| Constant | 0.0308 | 0.0423 | 0.0275 | 0.0847* | 0.0184 |
| | (0.0226) | (0.0421) | (0.0409) | (0.0442) | (0.0425) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 7397 | 7397 | 7397 | 7397 | 7397 |

Dependent variable: any member of the caste listed. SEs clustered by district in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A8: Intensive margin with rice price

| | (1) Brahman | (2) Kshatriya | (3) Ahir | (4) Chamar | (5) Muslim |
|---|---|---|---|---|---|
| log, Rice price, Rs/maund | 0.853$^{***}$ | -0.0124 | 0.764$^{***}$ | 0.663$^{***}$ | 0.221$^{*}$ |
| | (0.197) | (0.132) | (0.227) | (0.245) | (0.134) |
| log, Wage Rs/month | -0.117 | 0.110 | -0.506$^{**}$ | -0.748$^{***}$ | -0.228$^{*}$ |
| | (0.185) | (0.164) | (0.201) | (0.256) | (0.126) |
| Volatility, Rice price, Rs/maund | 0.0931 | 0.102 | -0.217 | 0.0354 | 0.170 |
| | (0.315) | (0.185) | (0.365) | (0.568) | (0.210) |
| Constant | -1.148$^{*}$ | 0.168 | 0.124 | 0.563 | 0.379 |
| | (0.676) | (0.341) | (0.444) | (0.580) | (0.308) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 874 | 2313 | 1593 | 1317 | 2142 |

Dependent variable: . SEs clustered by state in parentheses.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table I.A9: Intensive margin with rice price

| | (1) Brahman | (2) Kshatriya | (3) Ahir | (4) Chamar | (5) Muslim |
|---|---|---|---|---|---|
| log, Rice price, Rs/maund | 0.514*** | 0.164 | 1.090*** | 1.345*** | 0.718*** |
| | (0.178) | (0.182) | (0.239) | (0.372) | (0.170) |
| log, Wage Rs/month | 0.0264 | -0.0371 | -0.533* | -0.717 | -0.444*** |
| | (0.201) | (0.149) | (0.280) | (0.484) | (0.158) |
| Volatility, Rice price, Rs/maund | -0.761 | 0.00705 | 0.146 | 1.096 | -0.203 |
| | (0.595) | (0.333) | (0.599) | (0.901) | (0.355) |
| Constant | -1.084** | 0.0625 | -0.526 | -1.163 | -0.0664 |
| | (0.494) | (0.337) | (0.588) | (0.801) | (0.369) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 601 | 1246 | 915 | 777 | 1427 |

Dependent variable: . SEs clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A10: Intensive margin with rice price

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Brahman | Kshatriya | Ahir | Chamar | Muslim |
| log, Rice price, Rs/maund | -0.164 | 0.394* | 0.482* | 1.788*** | 0.562* |
| | (0.174) | (0.205) | (0.291) | (0.393) | (0.305) |
| log, Wage Rs/month | -0.0282 | 0.0211 | -0.482 | -0.670 | -0.0109 |
| | (0.286) | (0.253) | (0.344) | (0.578) | (0.184) |
| Volatility, Rice price, Rs/maund | 0.289 | 0.307 | -1.250 | -0.844 | 0.465 |
| | (0.704) | (0.493) | (0.827) | (1.081) | (0.682) |
| Constant | 0.249 | -0.385 | 0.397 | -1.493* | -0.566 |
| | (0.454) | (0.510) | (0.703) | (0.809) | (0.783) |
| Year FEs | Y | Y | Y | Y | Y |
| Observations | 230 | 523 | 441 | 483 | 628 |

Dependent variable: . SEs clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A11: Multinomial logit: caste shares

| | Male | | Female | | Children | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Brahman | | | | | | |
| log, Wage Rs/month | -1.490*** | -0.101 | -1.128* | 0.344 | -1.878** | 0.396 |
| | (0.481) | (0.355) | (0.622) | (0.608) | (0.827) | (0.905) |
| log, Colony wage Rs | -1.423*** | 2.784*** | -1.291*** | 1.074** | -1.265*** | 1.642* |
| | (0.173) | (0.818) | (0.212) | (0.544) | (0.432) | (0.852) |
| log, Rice price, Rs/maund | -0.396*** | -0.558 | -0.883*** | -1.611*** | -0.657** | -2.212*** |
| | (0.0952) | (0.382) | (0.225) | (0.375) | (0.268) | (0.671) |
| Volatility, Rice price, Rs/maund | 1.377*** | 0.340 | -0.205 | -1.311 | 0.836 | -0.909 |
| | (0.532) | (0.822) | (0.739) | (1.263) | (1.201) | (1.669) |
| Kshatriya | | | | | | |
| log, Wage Rs/month | 0.319 | 0.114 | 0.817 | 0.523 | 1.062 | 1.105 |
| | (0.247) | (0.276) | (0.555) | (0.507) | (0.766) | (0.719) |
| log, Colony wage Rs | 0.221 | -0.0227 | 0.558** | 0.735* | 0.785* | 1.127* |
| | (0.151) | (0.275) | (0.283) | (0.442) | (0.401) | (0.600) |
| log, Rice price, Rs/maund | -0.895*** | -0.915*** | -1.073*** | -1.656*** | -0.512** | -2.122*** |
| | (0.126) | (0.171) | (0.195) | (0.347) | (0.254) | (0.522) |
| Volatility, Rice price, Rs/maund | 1.099*** | 0.158 | 2.094*** | 0.707 | 2.780** | 0.617 |
| | (0.343) | (0.519) | (0.645) | (0.868) | (1.287) | (1.623) |
| Ahir | | | | | | |
| log, Wage Rs/month | 0.325 | -0.00111 | 0.643 | 0.294 | 0.119 | -0.360 |
| | (0.209) | (0.189) | (0.531) | (0.402) | (0.710) | (0.572) |
| log, Colony wage Rs | -0.00624 | -0.273 | 0.507* | 0.679** | 0.401 | -0.198 |
| | (0.180) | (0.237) | (0.295) | (0.337) | (0.393) | (0.802) |
| log, Rice price, Rs/maund | -0.144* | -0.126 | -0.425*** | -0.643** | -0.142 | -1.326*** |
| | (0.0793) | (0.156) | (0.152) | (0.287) | (0.221) | (0.446) |
| Volatility, Rice price, Rs/maund | 0.754** | 0.399 | 1.996*** | 0.725 | 1.558** | 0.466 |
| | (0.320) | (0.492) | (0.527) | (0.667) | (0.605) | (1.326) |
| Muslim | | | | | | |
| log, Wage Rs/month | -0.0722 | 0.228 | 0.145 | 0.249 | 0.198 | 0.391 |
| | (0.284) | (0.313) | (0.512) | (0.425) | (0.663) | (0.651) |
| log, Colony wage Rs | -0.376*** | 1.245*** | -0.415** | 0.851** | -0.155 | 1.960*** |
| | (0.144) | (0.333) | (0.187) | (0.355) | (0.271) | (0.514) |
| log, Rice price, Rs/maund | -0.429*** | -0.288 | -0.636*** | -0.669** | -0.427* | -0.502 |
| | (0.0962) | (0.190) | (0.138) | (0.263) | (0.232) | (0.417) |
| Volatility, Rice price, Rs/maund | 1.904*** | 0.104 | 2.265*** | 0.105 | 2.912*** | -0.171 |
| | (0.352) | (0.488) | (0.502) | (0.794) | (0.611) | (1.259) |
| District FEs | Y | Y | Y | Y | Y | Y |
| Colony FEs | Y | Y | Y | Y | Y | Y |
| Yr FEs | N | Y | N | Y | N | Y |
| Observations | 75153 | 75153 | 30982 | 30982 | 12376 | 12376 |

Chamar is the omitted caste. SEs clustered by district in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I.A12: Survival in Fiji, failure=exit

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Female | -0.0998*** | 0.0612 | -0.200*** | -0.289*** | -0.125** |
| | (0.0302) | (0.0495) | (0.0385) | (0.0601) | (0.0510) |
| Age | 0.0317*** | 0.0512*** | 0.0235 | 0.00181 | 0.0421 |
| | (0.00355) | (0.00543) | (0.0183) | (0.0269) | (0.0259) |
| $\text{Age}^2$ | -0.0000684* | -0.000176*** | -0.0000317 | 0.000411 | -0.000445 |
| | (0.0000374) | (0.0000425) | (0.000345) | (0.000498) | (0.000496) |
| Height (in) | -0.0360 | -0.0758 | -0.0895* | -0.128*** | 0.215 |
| | (0.0325) | (0.0610) | (0.0472) | (0.0417) | (0.145) |
| $\text{Height}^2$ | 0.000290 | 0.000439 | 0.000812** | 0.00108*** | -0.00159 |
| | (0.000249) | (0.000474) | (0.000364) | (0.000300) | (0.00115) |
| log, Rice price, Rs/maund | 0.117 | -0.0887 | 0.214** | 0.0608 | 0.350** |
| | (0.0844) | (0.140) | (0.109) | (0.183) | (0.139) |
| Volatility, Rice price, Rs/maund | -0.00995 | -0.0558 | 0.0256 | -0.291 | 0.258 |
| | (0.209) | (0.323) | (0.278) | (0.391) | (0.390) |
| log, Wage Rs/month | -0.0935 | 0.198 | -0.277** | -0.325 | -0.175 |
| | (0.106) | (0.183) | (0.133) | (0.210) | (0.174) |
| Ahir | 0.130*** | 0.108 | 0.141** | 0.261*** | 0.0603 |
| | (0.0430) | (0.0693) | (0.0551) | (0.0866) | (0.0718) |
| Brahman | 0.337*** | 0.318*** | 0.363*** | 0.517*** | 0.217** |
| | (0.0584) | (0.0965) | (0.0738) | (0.107) | (0.104) |
| Muslim | 0.156*** | -0.113* | 0.279*** | 0.209*** | 0.348*** |
| | (0.0397) | (0.0675) | (0.0497) | (0.0806) | (0.0633) |
| Kshatriya | 0.285*** | 0.174*** | 0.339*** | 0.430*** | 0.294*** |
| | (0.0398) | (0.0644) | (0.0508) | (0.0804) | (0.0660) |
| Observations | 22096 | 22096 | 18136 | 18136 | 14966 |
| Time | All | First 5 | After 5 | 5-10 | After 10 |

Standard errors in parentheses. Other caste binaries and district FEs not reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

## CHAPTER 2

## A (Paid) Passage to India:

## Quantifying the value of upper-caste status

# 1 Introduction

Why caste persists to this day in India has simultaneously a trivial and complicated answer: for some people, having a caste status is valuable. But what is the value of a caste, particularly for upper-caste individuals? Put differently, how much are members of different castes willing to pay for that status?

Although concepts of caste are arguably fluid, caste by and large remains a hereditary characteristic.[1] High-caste individuals marry other high-caste individuals and beget high-caste children. Individuals do not purchase caste status or receive it on the basis of other traits. In India today, caste-based reservation policies have changed the benefits that accrue to lower castes. However, unequal historical and contemporary benefits for upper-caste persons explicitly underlie such affirmative-action policies. Whether this is a correct assumption today and what the value of caste has been remain open questions.

Because the benefits of caste are neither traded on the market nor easily measured, I turn to a context in which Indians migrated and lost the aggregate benefits of caste, including social standing and economic opportunities: Indian indentureship to Fiji in the late nineteenth and early twentieth centuries. In this case, explicit laws regulating labor-market outcomes equalized labor-market opportunities across castes. Additionally, the geographic dislocation from India reduced the value of membership in a caste-centered social network as well as the psychological upper-caste benefit of living in India. Overall, the salience of caste dropped within Fiji.

[1] For the sake of this argument, Muslims and Christians each constitute a "caste."

Against this backdrop, I use unique data on the census of Indian entrants, deaths, and return migrants to India from Fiji first used in Persaud (2016). I link individuals to observe the exact timing of entry into and exit (if relevant) from Fiji. I also utilize institutional details about labor contracts and variation across entrant castes, ticket prices, and time.

Clear inter-caste differences in return migration appear in Figure II.1. Mortality appears roughly consistent across castes. However, around half of all high-caste Brahmans returned to India within ten years, whereas just around a quarter of low-caste Chamars did. The high-low caste differences are borne out in return data. If castes were the same, then such a difference should not exist, particularly since the ten-year mark guaranteed government-paid passage to India.

The price paid for returning to India differed across years and enables me to estimate a lower bound on the willingness to pay to return to India by members of various castes (alternatively, a lower bound on the compensating differential for remaining in Fiji). The willingness to pay for passage back to India is a measure of gross benefits for the individual. Effectively, this is a lower bound on the value of a given caste. I calculate this lower bound at roughly 2.5 years' worth of gross unskilled labor income.

This paper connects with theory on discrimination. Becker (1957) lays out the canonical framework for discrimination and places the costs of discrimination on the discriminated-against group, on the discriminating employer, or on the discriminating public. However, workers themselves may pay to discriminate. Here, firms do *not* discriminate in one market but do (or may) discriminate in another. The privileged group in the non-discriminating market then pays a cost to enter the discriminating market. This differs from discrimination with search models because the willingness to discriminate is known ex ante. Instead, this resembles barriers to entry in high-rent occupations or industries with the major difference that only some individuals are able to enter based on non-economic traits. This last point connects my paper to discrimination.

Previous research on return migration has focused on different returns to capital at home and abroad, currency shocks that change wealth, and optimal life-cycle earnings (see especially Yang (2006)). However, discrimination and inter-group differences remain understudied. Heterogeneous types of migrants may return differently not because of explicit economic motivations or returns to capital but rather because of preferential labor market access.

Finally, this paper connects to newer research from outside the US and western Europe on intergenerational inequality. Shiue (2016) finds that kinship groups matter for marriage and status transmission in pre-modern China. Caste, a kinship tie, has been used to explain India's relative backwardness. I can measure and estimate the economic role of caste explicitly in this setting in order to join this approach with other traditional methods of studying inter-group differentials.

## 2 Conceptual Framework

I build off Borjas (1987) with two types of workers. Borjas differentiates workers with a term capturing the skill premium in a given market. Here, I allow for differential skills but define inter-group differences as non-market, non-wage differences. I introduce a single market, denoted market 0, and two types, $\theta_0$ and $\theta_1$, who are distinguished only by observable non-economic characteristics. In the absence of discrimination, a member of either group receives a wage draw $\ln w_{i,0} = \mu_0 + \epsilon_{i,0}$, with $\epsilon_{i,0} \sim N(0, \sigma_0^2)$.

Suppose now there is an additional market, market 1, in which $\theta_1$ types receive an extra flow payment $\delta > 0$. This is a gross payment: it aggregates a wage differential, social status, and an amenity value of the location for type $\theta_1$. In contrast, $\theta_0$ types receive no additional payments. The net wage[2] for an individual $i$ in this market is then $\ln w_{i,1} = \mu_1 + \mathbb{1}[\theta_1]\delta + \epsilon_{i,1}$, with $\epsilon_{i,1} \sim N(0, \sigma_1^2)$.

The conditions for individual $i$ moving from market 0 to market 1[3] in a static migration model are

$$\begin{aligned} \theta_0: \quad \mu_1 + \epsilon_{i,1} - C &\geq \mu_0 + \epsilon_{i,0} \\ \theta_1: \quad \mu_1 + \epsilon_{i,1} + \delta - C &\geq \mu_0 + \epsilon_{i,0} \end{aligned} \tag{1}$$

where $C$ is the transportation cost to migrate. Figure II.2 shows this decision graphically with errors across markets distributed independently. The solid blue line is the 45° line emanating from the origin and represents indifference for type $\theta_0$. The additive scalar $\delta$ shifts down the dashed maroon line for type $\theta_1$.

In the aggregate, this implies that type $\theta_1$ individuals are more likely to migrate at a given level of $C$ because their net benefits exceed those of type $\theta_0$. Furthermore, for individuals with the

[2] This makes the generic assumption that utility is additively separable for $\theta_1$ types in $w$ and $\delta$.

[3] Like Borjas (1987), the reverse migration case is symmetric with a negative cost to migrating for $\theta_1$ types.

same $\epsilon$ draw but different types, the $\theta_1$ individual migrates for higher values of $C$ even when the $\theta_0$ individual does not.

The correlation $\rho$ between $\epsilon_0$ and $\epsilon_1$ only partly determines selection because $\delta$ mitigates the effect by lowering the threshold for $\theta_1$ types. Regardless of the value of $\rho$, $\theta_1$ types migrate more so $E[\epsilon_i|migrate, \theta_1] < E[\epsilon_i|migrate, \theta_0]$. If $\delta$ is sufficiently large, in the absence of general equilibrium effects on wages there can be complete migration across markets by type $\theta_1$ types.

This model illustrates how discrimination affects migration flows. Market inefficiencies in setting wages affect migration, so the observed level of migration from market 0 to 1 exceeds the economically efficient level.

## 3 Background

Indian indentureship in the British Empire began in 1838 as a response to the final emancipation of slaves in the British Empire. Mauritius, followed by West Indian colonies, pioneered the importation of Indian laborers brought in under long-term contracts. Until indentureship ended in 1917, well over one million Indians had been recruited for colonies all over the world.

Indian indentureship to Fiji began in 1879, several years after Fiji's annexation in the British Empire. Due to demographic crises and concerns about the impact of colonial firms on local workers, the colony turned to India to provide laborers for the infant sugar industry. Recruitment begin first in north India and was legalized from south India by the early 20th century.

Over the next 37 years, over 60,000 Indians were recruited and sent to sugar estates in Fiji. Figure II.3 shows Indians' districts of origin. Around 75% came via Calcutta from the United Provinces and the western part of the Bengal Presidency (later the state of Bihar). Large numbers also came from the coastal regions of the Madras Presidency in the south via Madras.

Unlike contemporaneous free migration to the Americas or private long-term contractual labor within British India, the colonial government in India organized and oversaw a monopsony of foreign labor destined for foreign colonies. The governments in recipient destinations applied for permission to recruit in India after passing local legislation governing Indian labor contracts. Colonial officials in India evaluated the applications and had to accept the terms prior to recruitment.

Because of this particular administrative set-up, the British colonial government in Fiji set wages by law, which were ratified by the government in India. Crucially for identification purposes, labor contracts specified ex ante the wage bill for the entire contract duration and the wages did not vary across caste. Unlike colonial English indentureship two centuries earlier, the contracts were flat across worker characteristics except for age; adults were paid more than children. This provision equalized wage opportunities for all adults. This also means that, in theory, workers would earn the same amount of wages during their indentureship periods and could have had the same stock of money at the conclusion of indentureship.

Figure II.4 shows the selection model outlined above graphically. The x-axis gives a theoretical distribution of wages in India and Fiji with the flattened Fijian wage structure represented by a horizontal line. The solid blue 45° line shows the migration decision for a low-caste person and the dashed red line shows the corresponding decision for a high-caste person.

The figure shows selection on wage draws in India. Given common support over the errors and correlations over time consistent with Borjas' formulation, high-caste persons are negatively selected relative to low-caste persons: $E[\epsilon_i|k_i = High] < E[\epsilon_i|k_i \neq High]$ where $k_i$ indexes caste. The lower wage returns for high-caste persons thus should have dampened their levels of return migration vis-a-vis low-caste persons, who had higher expected skills, in the absence of a $\delta$ effect. Alternatively, if the value of $\delta$ for high-caste persons had a distribution, the high-caste persons who did migrate had lower values of their status. In both cases, marginal upper-caste persons migrated. The different interpretation of initial selection only helps, since it either gives the true $\delta$ or a lower bound of $\delta$ in India across all upper-caste persons.

A second unique feature of Indian indentureship was the explicit provision of return migration. In the case of Fiji, individuals were allowed to purchase their own passage back after completing a five-year indentureship term. After ten years of continuous residence, individuals received a paid passage back to India provided by the colony.

Bureaucrats in Fiji negotiated for return passage back for a given ship heading to India. The legislation authorizing Indian indentureship to Fiji required bureaucrats to arrange transportation back to India for those desiring it. Even in the Indian famine year of 1896-7, when the Protector of Immigrants noted that Indians would be better off remaining in the colony, transportation was still arranged to return paying passengers and so-called 'ten-years' people' eligible for a free passage.

This price was negotiated *before* Indians made a decision to return and was exogenous to their individual characteristics. The shipping line, too, did not know who would choose to return to India so the passage costs did not vary based on the actual number of returners. The colonial government ensured a minimum payment to the ship in the event of a small number of emigrants. In the one case when too few emigrants desiried to go back to India, the colonial government paid a penalty to cancel the standing contract and instead negotiated a new contract even though it would have been more cost-effective to delay repatriation until the following year.

After negotiating, bureaucrats then notified free Indians in the colony of the ship's itinerary, price, and details of the voyage. Therefore, Indians at a given point in time faced different costs to return: infinite during the first five years, positive after five but before ten, and zero after ten years. The shipping line earned a constant marginal revenue from each additional returner.

Repatriation during the first five years was extremely rare and occurred only for extreme medical causes such as leprosy. Any indentured Indian repatriated for medical reasons had to pass a medical inspection, which screened out attempts at breaking contract and gaining a free passage back to India. A special case, ejection at the start of indentureship, occurred due to errors in recruiting, such as failing to identify a communicable disease, or extreme frailty due to the voyage.

The caste backgrounds of Indians going to Fiji were varied and reflected the composition of castes back in India. The main castes of interest here are from north India due to the larger number of migrants and ease in identifying castes in transcribed data. I select several main groups: Brahmans, Kshatriyas[4], Ahirs, Chamars, Muslims, and others. Muslims, while not a caste, do form a significant and distinct group in the data. For the remainder, Brahmans were a high, priestly caste. Kshatriyas were similarly a high caste, albeit lower than Brahmans, with traditional landowning and warrior roles. Ahirs were a middling agricultural caste. Chamars were a lower caste and worked as laborers, particularly as leather workers. Finally, I combine the remainder into a larger "other" group.

One particular note of interest for Brahmans is the high psychological or social cost of leaving India and becoming ritually impure (Hugh (1974) 46, Mangru (1987) 57). These costs were incurred for having crossed the ocean ("kala pani," literally black water, in Hindi). The British East India Company explicitly exempted Brahmans from overseas service (e.g., in China in the Opium Wars) to prevent discontent or mutiny.This psychological cost could have translated into real costs of

[4] Kshatriya here includes Chattris, Jats, Rajputs, and Thakurs.

purification rites. The discrete ocean-crossing meant that anyone who left, regardless of the time abroad, would have to pay the purification costs. Thus, the value of $\delta$ for Brahmans was lower or the cost to return higher, both of which lowered the incentives for Brahmans to return.

# 4 Data and empirics

## 4.1 Data

The census of Indian entrants to Fiji from 1879 until recruiting ended in 1916 is housed on microfilm at the National Library of Australia, Canberra. All Indians were issued with an emigration pass in India at the time of departure. I digitized individual-level data from these emigration passes of all 60,000-plus indentured immigrants. Due to the bureaucratic nature of the passes, the information contained therein should truly reflect the leaving population. Bureaucrats had no incentive to manipulate responses, as the passes were used for adjudication of contract disputes. The Indian indentured workers themselves also had incentives to tell the truth because passes were used to identify them for repatriation to India, collect other benefits, and remit estates to India in the event of death abroad.

To prevent fraud, officials collected information on self-reported background, such as next of kin, and from observation, notably physical marks and height. Although some bodily marks could be common, such as scars on the knee or a pockmarked face, many others were not. Moles, tattoos, birthmarks, and physical disabilities were all noted and would have distinguished individuals in addition to the names and familial information. I have discovered no cases of death or return records updated with notes about fraud and mistaken identity, which could have happened if a living person's identity was stolen.

I digitized the census of return migrants and deaths from 1879 through 1913. These, too, are stored on microfilm at the National Library of Australia. Individuals received a unique numeric identifier on entry into Fiji that bureaucrats used in all official business. I am thus able to bypass problems with string matching by using the true numeric identifier to link individuals to death and return migration records.

Furthermore, except for the earliest years, the return migration records indicate the reason for repatriation. I focus here on returners via Section 113 P, proceeding under passport. These individuals explicitly paid for their passage. Because the records were kept as official government documents and were used to prevent false claims for return passage, they are assumed correct.

All entrants in 1908 and earlier, who were eligible for paid passage back to India, form the analytical sample. This captures entrants who were eligible to return to India before World War I. The war disrupted non-military shipping in the Indian Ocean and so comparisons in years before, during, and after are invalid.

The value of $\delta_k$ for a given caste $k$ entered into the migration decision of an individual when she made a decision to remain or leave. The easiest point to identify $\delta$ is at the end of the indentureship contract. This bypasses problems with multiple periods and having to assign a (potentially caste-varying) discount factor to weight time in Fiji. Therefore, I focus on the one-year window after the fifth year, i.e., after the indentureship ended, as the relevant period of return migration.

I take the negotiated passenger rates from Fiji's *Annual Reports* from 1892-1913, since these years have the explicit reason for repatriation and ship price data. Figure II.5 shows a scatter of the prices in pounds and the percentage of newly emancipated indentured servants who chose to pay for repatriation at the first possible time.[5] Ship prices show significant variation over time, from a high of £10 at the beginning of the period to a low of £4 by the end. Repatriation varies, too, over the period. The scatter of ship prices and return patterns in Figure II.6 gives suggestive evidence of some economic responses to prices. Higher prices were generally accompanied by lower repatriation rates.

I construct two measures of prices for those who did not return to India. First, I take the average of ship prices over years 4-6. This enables me to deal with fuzziness about the exact start and end dates of indentureship. Second, I take the exact price for the first shipping season after an individual's indentureship ended. For example, for a person who entered in January 1891, I find the first ship after January 1896 and take that price. In years with ambiguity due to the proximity of voyages and differing prices, I use the average price of the passage from the possible ships.

I restrict the sample to adults at the time of initial departure from India who reached Fiji alive, since they consciously made choices to sign indentureship contracts and faced the same conditions there. Children also face potentially different returns to repatriation if caste benefits accrue more to adults, e.g., through labor-market opportunities.

[5]Individuals who died or were involuntarily repatriated are excluded from the analysis.

## 4.2 Empirics

The adult Indians who travelled to Fiji were mostly men in the prime of life. They came from a variety of caste backgrounds and districts. Table II.1 shows basic summary statistics at the time of entry into Fiji. Almost three-quarters of them came from the United Provinces (modern-day Uttar Pradesh). One key point is that my five chosen caste groups (Brahmans, Kshatriyas, Ahirs, Chamars, and Muslims) comprised over half of all north Indian entrants. This helps in identifying differences across the groups.

The fate of the arrivals over the first five years during indentureship and just after, in years five - six, form the crux of my analysis. I show summary statistics for these periods in Table II.2. Just like Figure II.1 shows for all types of repatriation, there were clear differences across caste in paid return migration in the first period after indentureship ends. Mortality may have been higher for higher castes but involuntary repatriation, including rejection, remained low for all castes. Higher proportions of Brahmans and Kshatriyas returned to India than low-caste Chamars.

Figure II.7 shows the time trends for mortality and repatriation. Rejection spiked at the start and then dropped off as only fit persons remained. Similarly, repatriation for incompetent and incapable persons, which included medical reasons, grew in the first five years and dropped off. Higher mortality at the start of the indentureship period occurred due to “seasoning” in the new destination and among ill new arrivals. Mortality increased at the start of indentureship and then fell off.

The basic set-up is a probability model of the form

$$p = P(Return_{ijkt}) = \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 Female + BI_k + \Gamma I_j + \delta price_t + \epsilon_{ijkt} \quad (2)$$

where $i$ is the person, $j$ is the district of origin, $t$ is the year, and $k$ is a caste. $B_1$ is a vector of coefficients on caste dummy variables. For a caste $k$, the corresponding coefficient is $\beta_k$. I omit Chamars, the low caste. I omit year fixed effects because, with only a few exceptions, the negotiated ship price did not vary within a year. For some years, only one ship sailed back to India. For others, the same price was negotiated for all ships going back.

I use the latent utility representation from Equation 2 to back out a relative economic value of a caste $k$. I take the total derivative and hold the other variables and utility constant:

$$dp = 0 = \beta_k I_k + \delta d(price) \Leftrightarrow \frac{d(price)}{I_k} = \frac{\beta_k}{-\delta} \tag{3}$$

Demand curves should slope down so $\delta \overset{!}{<} 0$. Figure II.5 generally shows this pattern. If this is the case, then the ratio of the coefficients represents the relative value of caste $k$ to Chamars in pounds sterling, the currency used to price ship passages.

Table II.3 shows the main specification with a linear probability model. Column 1 shows the base model from the equation. The coefficients on the castes are positive and significant at the 1% level. Furthermore, the point estimate on the Brahman coefficient is greater than the Kshatriya coefficient, which is in turn greater than the others. The difference between Brahman and Kshatriya is not significant, though.

As expected, $\hat{\delta}$, the coefficient on price, is negative; here, the demand curve for return passage to India does slope down. I then generate a value for each caste per Figure 3 based on its coefficient and the price coefficient. Brahmans valued return to India at least £10 more than Chamars, whereas Kshatriyas valued it just under £6 more than Chamars.

In column 2, I trace the marital status of individuals at the time of departure from India. I distinguish between married to a spouse who remained in India and married with an accompanying indentured servant. The effect of travelling with one's spouse is ambiguous: while it doubled potential earning power, it also signalled willingness to migrate, maybe even permanently. On the other hand, leaving a spouse behind in India probably signalled willingness to return to India. While I do not have information on subsequent marriages or deaths of absentee spouses, this initial type of selection may have borne out differential return migration five years later.

As expected, the coefficient on being married to a spouse in India is positive and significant at the 5% level. The coefficients on the castes, though, are effectively unchanged from column 1. The calculated caste values drop slightly across the board. In terms of magnitudes, the coefficient on Brahman is double that of the coefficient on being married to a spouse in India.

Because females return at significantly lower rates than men and there are potentially differential opportunities and caste-related payoffs for men and women, I modify the regression equation by adding caste-female interaction terms. Column 3 of Table II.3 shows these results. The coefficients on Brahman and Kshatriya both increase whereas the coefficients on Muslim and non-Chamar, other are effectively unchanged. All remain significant at the 1% level.

Interestingly, the female interaction terms point to stark differences between men and women in the value of returning to India. Tests of equality between both Brahmans and Kshatriyas and the omitted group, Chamars, yield the same finding: women of *all* castes returned at the same rate. Caste benefits may have accrued to men, not to women, or certain local benefits may have accrued only to women that tipped their decisions towards remaining. Widow remarriage may have been one, for instance, given the overwhelmingly male migration.

Regardless of the reasons, the coefficient on Brahman rises significantly in comparison to columns 1-2 while the coefficients for others remain mostly similar. The lower-bound estimated value for a Brahman in India rises to roughly £14 and for a Kshatriya to over £7 relative to a Chamar. In column 4, which includes the marriage variables, the values for the castes remain largely unchanged.

The ordering in all the columns, which is really a male value, reflects the hierarchy presented above. The value for Brahmans exceeds the value for Kshatriyas exceeds the value for Ahirs exceeds the value for Chamars. The catch-all group for other non-Chamars shows a higher value than Ahirs but lower than Kshatriyas. This reflects a pooling of multiple castes.

To put the value for male Brahmans into perspective, it is useful to compare against rupee-denominated wages in India. The sterling-rupee exchange rate was fixed at 15 rupees per pound in 1898. In the decade building up to this, the rupee traded at a weaker rate and fluctuated from 15 - 18 rupees per pound. For the most conservative calculation, I take 15 rupees per pound as the exchange rate.

The average unskilled wage in north India did not exceed five rupees per month in the late nineteenth and early twentieth centuries. Using this and a large value of the annual discount factor ($\beta \approx 1$), I calculate the relative value of being a Brahman as at least 2.5 years worth of gross unskilled wages. This uses the estimate from column 1; using the estimate from column 4, the relative value of a Brahman goes up to 3.5 years. Similarly, the value of being a Kshatriya is around 1.5 years worth of gross unskilled wages. Varying the value of $\beta$ increases the value for being one of these upper castes relative to being a Chamar.

### 4.3 Robustness

The results above could be biased due to data problems or failures in my assumptions about caste. Several main potential issues arise including measurement error, omitted variable bias in the accounting identity used to generate the relative value of a caste, and endogeneity of ship prices.

First, the ship price used in Table II.3 takes an average over several years of prices in order to capture the possible ships that an individual could have taken. However, this could introduce classical measurement error if the averaged price deviates from the real price. To check for this, I re-run Equation 2 with the strict price taken from the next ship. Appendix Table II.A1 shows these results. The sample size drops slightly because individuals miss the cutoffs at the boundaries of the period. The estimated coefficients are higher than in Table II.3, but the calculated value of a caste is effectively unchanged.

Second, the estimated $\hat{\delta}$ measures a different value than the theoretical $\delta$ from Section 2. Here, instead of explicitly differencing off group means, I rely on an accounting identity that $w_{HighCaste,Fiji} \geq w_{LowCaste,Fiji}$. If this were not the case, it would reflect differences in skill that the flattened nature of the labor market in Fiji compensated. Put differently, if higher castes were relatively worse skilled, their lower wages in Fiji could have pushed them back to India. In this case, what I observe is not the value of a high caste but rather a negative correlation of $\epsilon$ terms for the high-caste individuals.

Two facts militate against this. The emigration passes indicate occupation. For almost all individuals, occupation is listed as "cultivator" rather a than skilled occupation. This cuts across caste; there is no caste-related variation in occupation. This reflects the underlying distribution of occupations reported in the Census of India. In 1911 in the United Provinces, for example, just under half of all male workers and around 40% of female workers were classified as "cultivators of all kinds," a category exclusive of landlords. The percentages for male workers rose to 75% for Brahmans and Ahirs and stayed near the provincial mean for Kshatriyas and Chamars. The difference between Chamar and Brahman could arise from this difference in cultivation. However, that would not explain the significant difference between Kshatriya and Chamar (both low percentages of cultivator) or the difference between Ahir and Brahman (both high percentages).

Additionally, the tasks during indentureship and after indentureship were by and large unskilled manual labor tasks in commercial agriculture. The skill premium would have been developed during the five years of indentureship that all Indians completed before being released. Rather than bringing caste-varying human capital from India, the conditions of Fiji required colony-specific, sugar-specific human capital gained in Fiji and accessible to all workers.

Mortality could have differed across castes, which would skew the valuation across of a return passage across castes. Figure II.8 does show differences by caste. However, these mortality trends are unadjusted. To account for differences across origin area, caste, and other demographic features, I run a Cox proportional hazard model for mortality during the first five years.

Table II.4 shows the results for the survival function. As opposed to Chamars, the omitted group, Brahmans do have slightly higher mortality but it is not significant. Additionally, the mortality survival figure (Figure II.8) does show a flattening of Brahman mortality starting around year four. Additionally, there could be yearly variation in mortality.

To examine this further, Figure II.9 shows the proportion of deaths by caste in the year prior.[6] Brahmans died at lower rates than Chamars until around 1905, and Kshatriya mortality tracks well with Chamar mortality. I also calculate lagged five-year mortality rates, i.e., mortality rates covering the period of indentureship.[7] Appendix Figure II.A1 shows these. Because this measure comes from the one-year lagged measure, the pattern of mortality looks similar: prior to around 1905, Brahmans exhibited lower mortality than Chamars. Interestingly, Kshatriya mortality exhibited an increase in the early 1900s even as mortality for the other two castes declined.

To account for differential mortality effects by caste, I first re-run the main specification for years through 1905 only in Table II.5. During this period of time, since high-caste persons died at lower rates than low-caste persons, the mortality channel should not affect induce higher return to India from high-caste persons. The sample size drops, which increases the bounds on the caste values. However, the value for Brahman in column 4, the preferred specification, remains around £8. The value for Kshatriyas has increased from the main Table II.3 to around £14. The direction and magnitudes for both castes are consistent with the earlier results. However, the drop in sample size increases the calculated standard errors greatly.

[6] I exclude returners to India in a given year from the denominator. This will have the effect of raising mortality for higher castes, who returned at higher rates.

[7] The numerator is the number of deaths of persons from $k$ caste from $t-5$ through $t-1$. The denominator is the sum of the number of persons of $k$ caste during this period. A person who appears in multiple years is counted multiple times. As before, I exclude returners to India in the year in which they left.

As an alternative robustness check with the larger sample, I include a covariate for caste-specific mortality in the prior year. Table II.6 shows the results of Equation 2 with an addition covariate for caste-specific mortality in the prior year. The calculated values for Brahman and Kshatriya return resemble those from the main specification. The value for Brahmans increases and the value for Kshatriyas decreases slightly in the preferred full specification (column 4). Mortality rates are not significant themselves, though, which implies that recent mortality may not have affected return to India.

I also check using the longer, five-year mortality rates in Appendix Table II.A2. The values for Brahman grow to around £14 and also increase for Kshatriyas in the full column 4. Here, mortality rates are positive and significant at the 10% level in odd columns and at the 5% level in the even columns. This is consistent with higher intra-caste mortality raising return migration. The economic significance is low, though; the mortality difference between Brahmans and Chamars of roughly 0.02 translates to roughly 1/8 of the effect of being a Brahman on returning.

A related criticism is that higher castes may have had higher returns to capital in India due to, for example, better access to land. In this case, the criticism fails to account for the fact that I calculate a gross measure of caste value that encompasses caste-linked higher flow utility in India. The existence of different capital markets or investment opportunities for different castes is precisely part of the value of membership in a particular caste. This defense also extends to credit constraints. The wages paid to workers in Fiji were sufficient for paying return passage. Even with inflated deductions for food, a worker could earn at least £9 per year during indentureship. This would be more than sufficient for an individual; for a one-parent household, this would still have supported return passage for the parent and four or more children. Direct consumption of wages in Fiji or sending remittances to India could have induced a credit constraint; however, such decisions would have been individually rational for lower-caste Indians.

More formally, it follows that the ratio of marginal utility of consumption to the marginal utility of saving was higher for lower castes than higher castes. If a return passage to India is part of the potential investment package for a person, $r_{High,India} \geq r_{Low,India}$. Credit constraints after five years come from utility-maximizing decisions over the lead-up period, and different capital returns reflect the benefits of high-caste status.

Third, ship prices could be endogenous. If better economic conditions in India affected ship prices and differentially the returns to being a high caste, then price may be highly correlated with caste. There may be omitted variable bias because of an external shock in India both driving up ship prices (reflecting opportunity cost) and return to India. However, a key strength of my identification strategy is that ship-level prices were negotiated ex ante by British officials without reference to the caste structure of the population. There was no price discrimination in ship prices by caste. At the time of departure, prices were exogenous to the individuals. Although the price was not dictated to the ship but negotiated, this mitigates some of the problem. Furthermore, if return of high-caste individuals and ship price were positively correlated, the calculated value for a caste would be biased downwards.

Fourth, the functional form choice may affect the results. I re-run the main specification using logit in Appendix Table II.A3. In all four columns, the values from the logit for Brahman fall to just under £6 and those for Kshatriya to just under £5. Similarly, I re-run the main specification using probit in in Appendix Table II.A4. The values here for Brahman and Kshatriya look similar to the logit in all four columns, with values for Brahmans hovering around £6 and for Kshatriyas just under £5. While both the logit and probit give lower estimates of the relative value of caste, the values themselves are the same order of magnitude as the linear probability model. They are also statistically higher than zero, i.e., there is a positive value relative to Chamars. The functional form may adjust the value, but the direction is still the same and shows a high value to being a high caste.

Fifth, the analysis above excludes both the possibility of paying for passage after year five but before year ten and the continuation value of remaining in the colony. Conceptually, this is a repeated series of choices for an individual until her death or return to India and is an analogue of the single-time choice at year five. I use the same population as above (eligible to return home starting in 1890), with some repeated observations for individuals who did not return immediately. Functionally, this is

$$p = P(Return_{ijkt}) = \beta_0 + BX + \Gamma_1 I_j + \Gamma_2 I_t + \Gamma_3 I_{TimeToFree,it} + \delta price_{it} + \epsilon_{ijkt} \tag{4}$$

where $X$ is the same demographic variables as above, with the added difference that age and squared age both vary at the year level. Here, because price now varies at the yearly level between those who must pay versus those who receive a free passage, I include time fixed effects. I also include several dummy variables for time until free passage (for years$= 1 \ldots 5$). The coefficients on these dummies capture the scale effect of the option value of staying under the assumption of linearly separable option values.

Table II.7 shows the results for the repeated choice. Like the earlier Table II.3, the coefficient on Brahman is positive and significant across specifications. The calculated value of a Brahman hovers around £6 for all specifications, and the Kshatriya value rises to a similar value between £5 and £6. The values for Muslim and other non-Chamar individuals also rise. The magnitudes are lower than the previous specification but still show a large, significant difference between upper castes and Chamars.

Sixth, there could have been ship capacity constraints going back to India. In this case, some Indians desiring to return to India would have been deferred until a later year. I plot the ratio of returners to India over last shipment from India in Figure II.10. In all but three years, the number of in-migrants exceeded the returners. Of these three, one return shipment was effectively the same size as the previous in-shipment. This ship, the Erne, was followed closely by another ship from India, which could have been used for return migrants but was not. This is evidence that there was not excess demand for return to India. Another ship, the Moy in 1893, brought far fewer passengers from India than normal before disembarking with a large number of returners. Adjusting the latter for the total carrying capacity either before or after 1893 lowers that ratio below one. After these adjustments, only one shipment could have caused problems with timing due to supply constraints back to India.

Seventh, height could have mattered more than caste identity. There is a long literature of height affecting economic outcomes and migration; see Bleakley, Costa, and Lleras-Muney (2013) for the general impacts and Kosack and Ward (2014) and Spitzer and Zimran (2014) for examples of height selection in migration. I re-run the main specification to include height in inches and a polynomial in height to capture this effect. In Table II.8, there is no change in the value of Brahmans or Kshatriyas. Both height and height squared are signed according to theory, but the results are not economically or statistically significant. In particular, the linear increase of a half foot, ignoring the negative quadratic, is roughly the same effect as being a Brahman.

# 5 Conclusion

The persistence of caste and entry into a caste-wised segmented labor market may explain these differences. While caste has changed since the late nineteenth century, it remains salient today. I offer some of the first quantitative evidence showing that caste translated into at least 2.5 years' worth of income. In a world near subsistence with improperly functioning credit markets and path dependence in land tenure (see Banerjee and Iyer (2005)), years' worth of income translated into large gains for upper castes.

High-caste individuals wanted to return to India rather than remain in Fiji more than low-caste individuals. In the aggregate, Brahmans and Kshatriyas were more willing to migrate during periods of paid passage and paid more for return passage. Holding the value of being an Indian in Fiji constant across castes, I estimate relative values for upper-caste Indians of returning to India that are consistent with large gains to belonging to an upper caste. The ordering, too, is consistent with priors on the social standing of castes: Brahmans were more valuable than Kshatriyas, who were more valuable than Ahirs, who were more valuable than Chamars.

Interestingly, the returns to caste are gendered. Only men exhibit differential return migration to India, and the relative value of a particular caste presented here is really the relative value for a male of that caste. One possibility is that the status of women, irrespective of caste, was higher in Fiji than India. Given the heavily skewed sex ratio in the colony, it is possible that women were able to raise their status. In this optimistic case, a woman's caste was immaterial, and there could have been caste intermarriage. In the pessimistic case, women in India were treated the same and caste-specific gains accrued only to men. The accounting identity here is flipped from the men's identity; differential return migration would have come from better or worse conditions in Fiji, not in India. The low levels of female return migration point to some version of the optimistic scenario, but the two are not mutually exclusive.

One potential threat to identification is the undercount of children and the lack of knowledge of marriages. Having more children would raise the cost of returning, which is currently unaccounted. Further work on caste exogamy will shed light on the sharp male-female split, since marrying "up" for women could have resulted in a new caste status. Research on marriages and births using undigitized records held in New Zealand will resolve this.

The slight modification to the Roy Model above to include a group-specific benefit outside of skill is consistent with the observed behavior and challenges a Borjas-style, human-capital argument for return migration. The flatness of the labor market in Fiji and the overwhelmingly agricultural economy back in India lend themselves to a Becker-style interpretation. Occupational structure does not explain the differences between castes with similar proportions of cultivators in the aggregate back in India; this weakens the human capital case. A schema in which Brahmans and Kshatriyas were much higher skilled would have to invert centuries of accepted beliefs on their upper-caste status in order to give an internally consistent migration selection under the Roy Model.

Potential differences in the returns to physical capital due to different investment opportunities or access to other capital sources would have to reflect an underlying economic difference, which could have come from inter-caste inequalities at the time of re-migration. Membership in an in-group could confer group benefits not related to discrimination per se but statistical discrimination. Purely psychological reasons for return migration may also generate the observed behavior. Differences in opportunities within a group and not related to productivity, i.e., Becker-style discrimination, is consistent with the data but with human capital as the driver of return migration.

# References

Banerjee, Abhijit and Lakshmi Iyer (2005). "History, institutions, and economic performance: the legacy of colonial land tenure systems in India". In: *The American economic review* 95.4, pp. 1190–1213.

Becker, Gary S (1957). *The economics of discrimination*. University of Chicago press.

Bleakley, Hoyt, Dora Costa, and Adriana Lleras-Muney (2013). *Health, Education and Income in the United States, 1820-2000*. Tech. rep. National Bureau of Economic Research.

Borjas, George J (1987). "Self-Selection and the Earnings of Immigrants". In: *The American Economic Review*, pp. 531–553.

Hugh, Tinker (1974). *A New System of Slavery: The Export of Indian Labour Overseas, 1830–1920*.

Kosack, Edward and Zachary Ward (2014). "Who Crossed the Border? Self-Selection of Mexican Migrants in the Early Twentieth Century". In: *The Journal of Economic History* 74.04, pp. 1015–1044.

Mangru, Basdeo (1987). *Benevolent Neutrality: Indian Government Policy and Labour Migration to British Guiana, 1854-1884*. Hansib Pub. Limited.

Persaud, Alexander (2016). *Risk mitigation and selection under forward contracts: 19th-century Indian indentureship.*

Shiue, Carol (2016). *Social Mobility in the Long Run: An Analysis with Five Linked Generations in China, 1300 1900.*

Spitzer, Yannay and Ariell Zimran (2014). "Migrant Self-Selection: Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907–1925". In: *Manuscript, Dept. Econ., Northwestern Univ.*

Yang, Dean (2006). "Why do migrants return to poor countries? Evidence from Philippine migrants' responses to exchange rate shocks". In: *The Review of Economics and Statistics* 88.4, pp. 715–735.

Figure II.1: Repatriation and mortality trends, pre-1904 entrants

Figure II.2: Migration decision

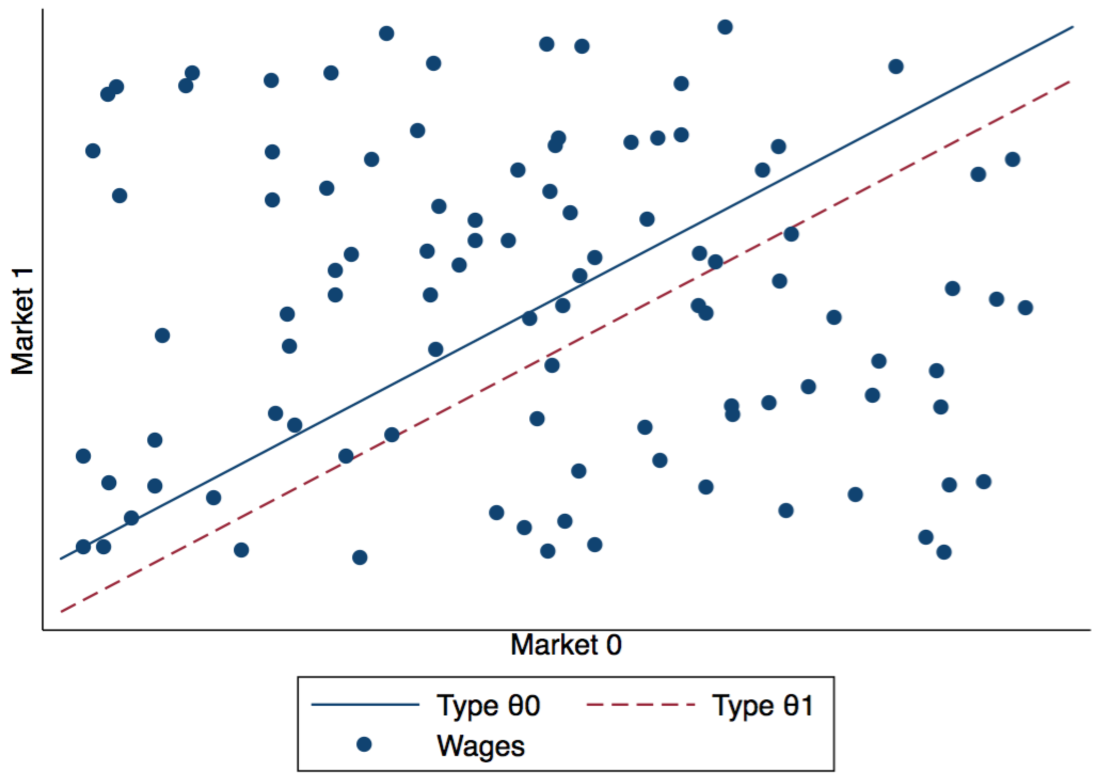


Figure II.3: Districts that sent Indians to Fiji

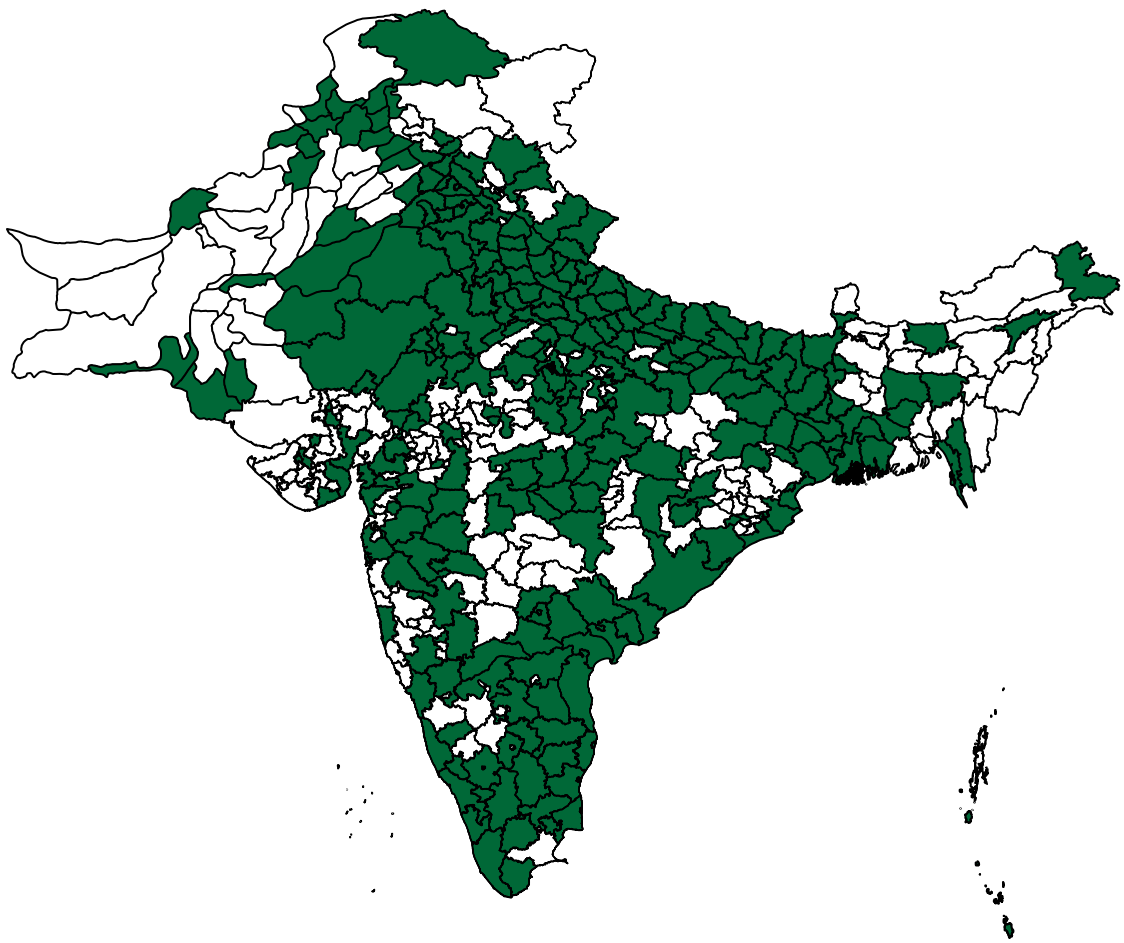


Figure II.4: Theoretical initial selection

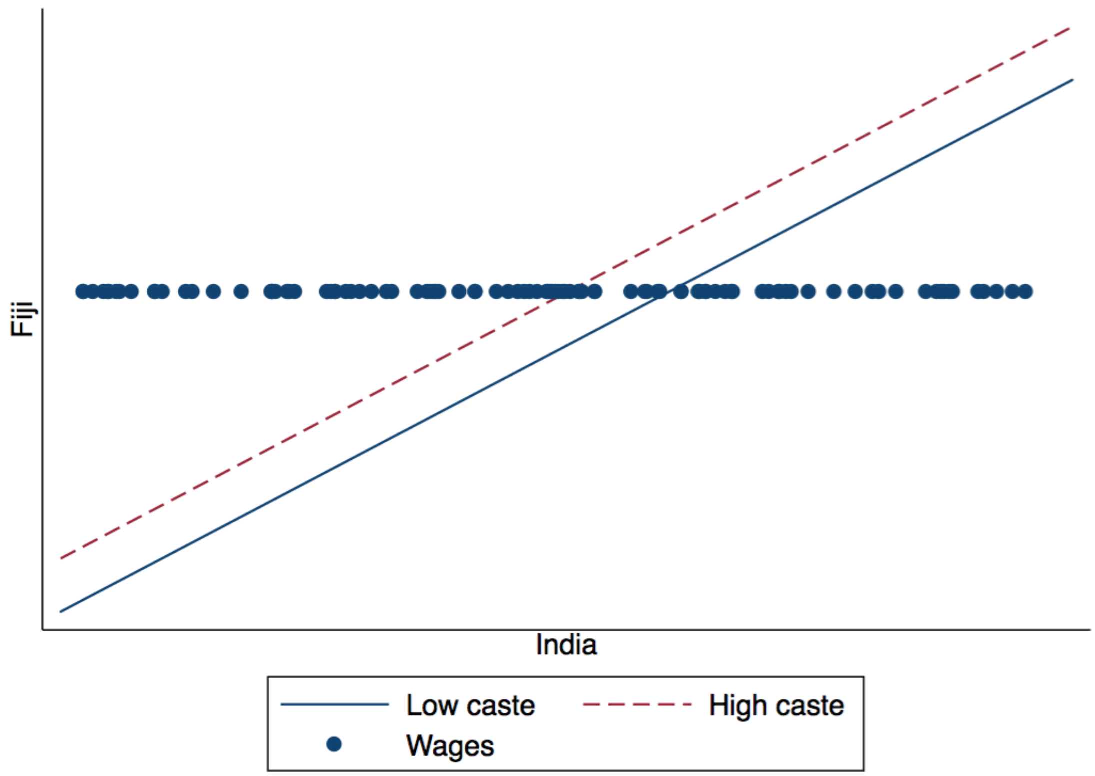


Figure II.5: Ship prices and return patterns

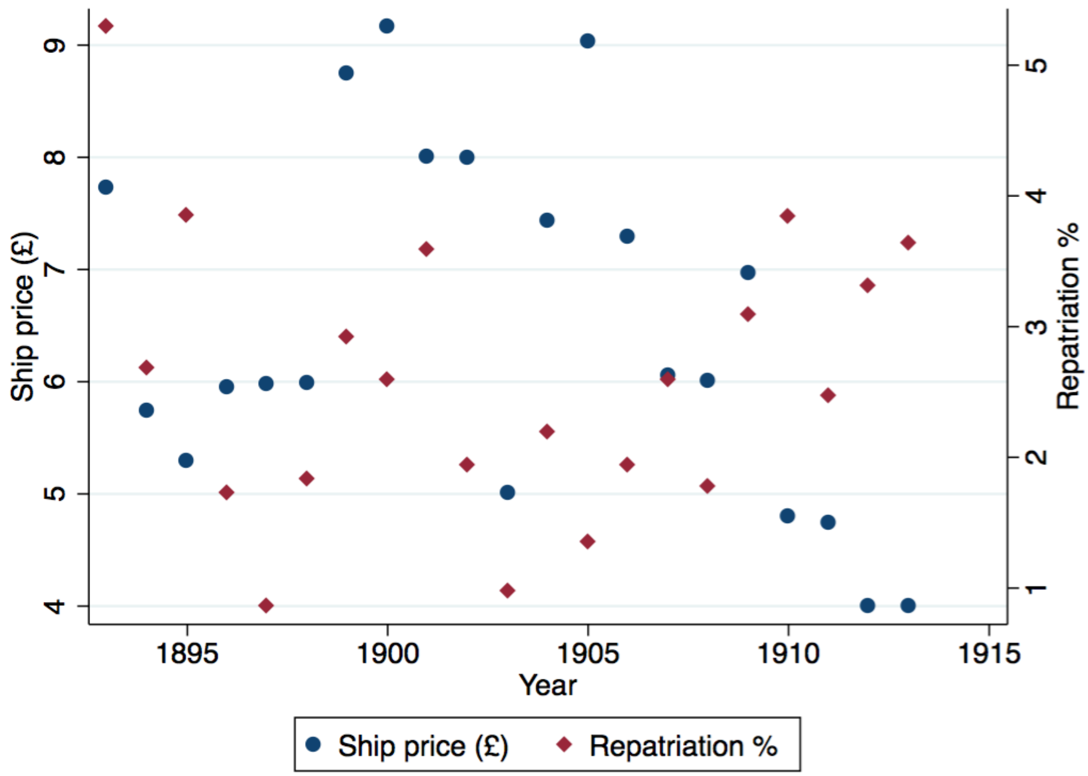


Figure II.6: Ship prices vs return patterns

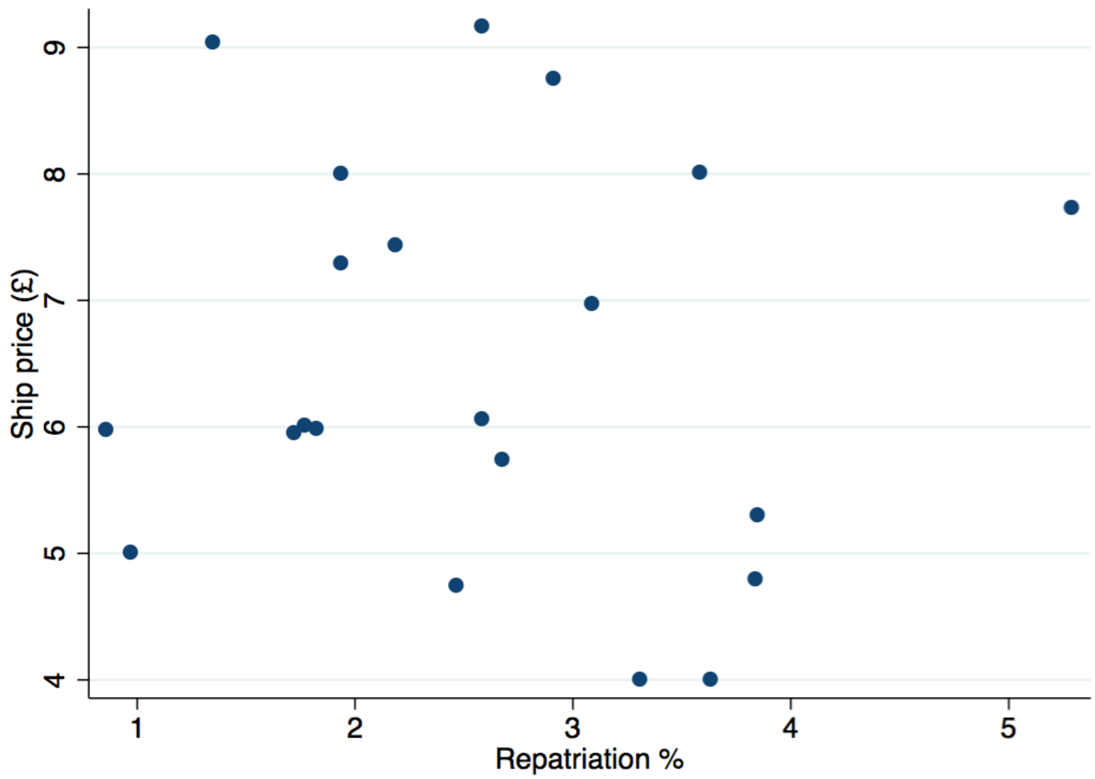


Figure II.7: Repatriation and mortality trends

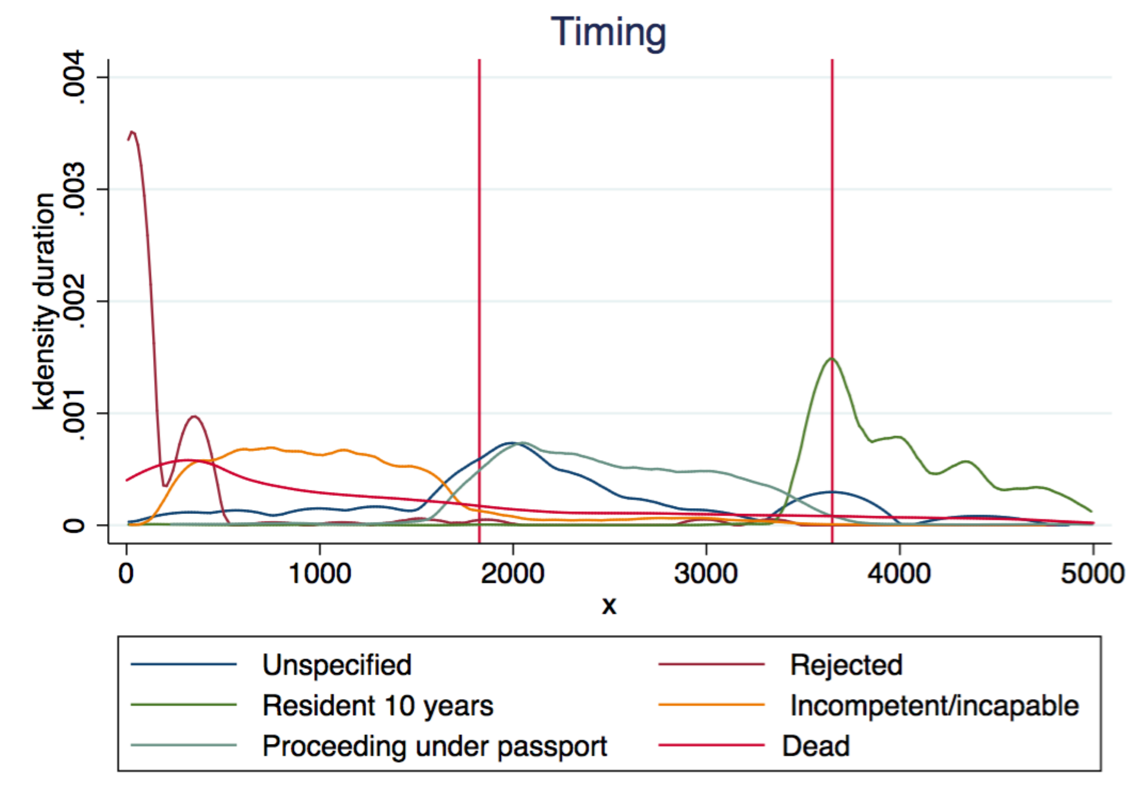


Figure II.8: Survival through year 6

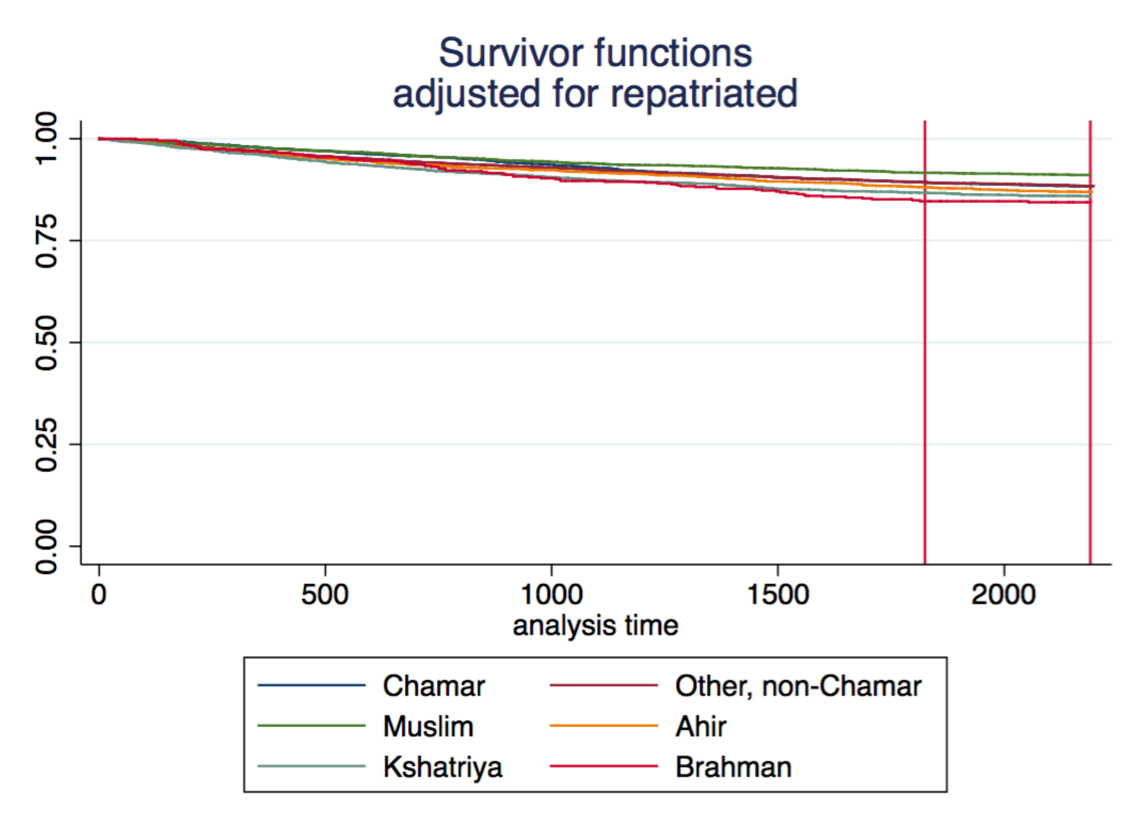


Figure II.9: Mortality prior year by caste

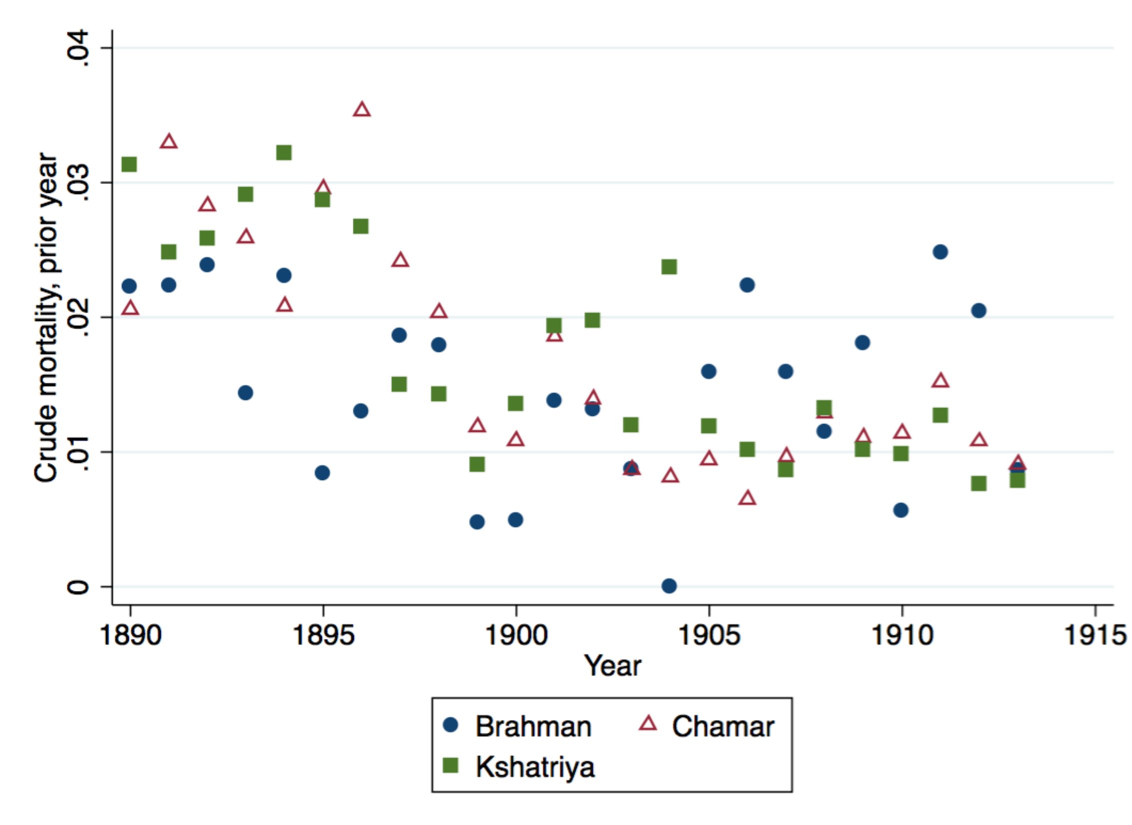


Figure II.10: Return shipments as a proportion of last shipment

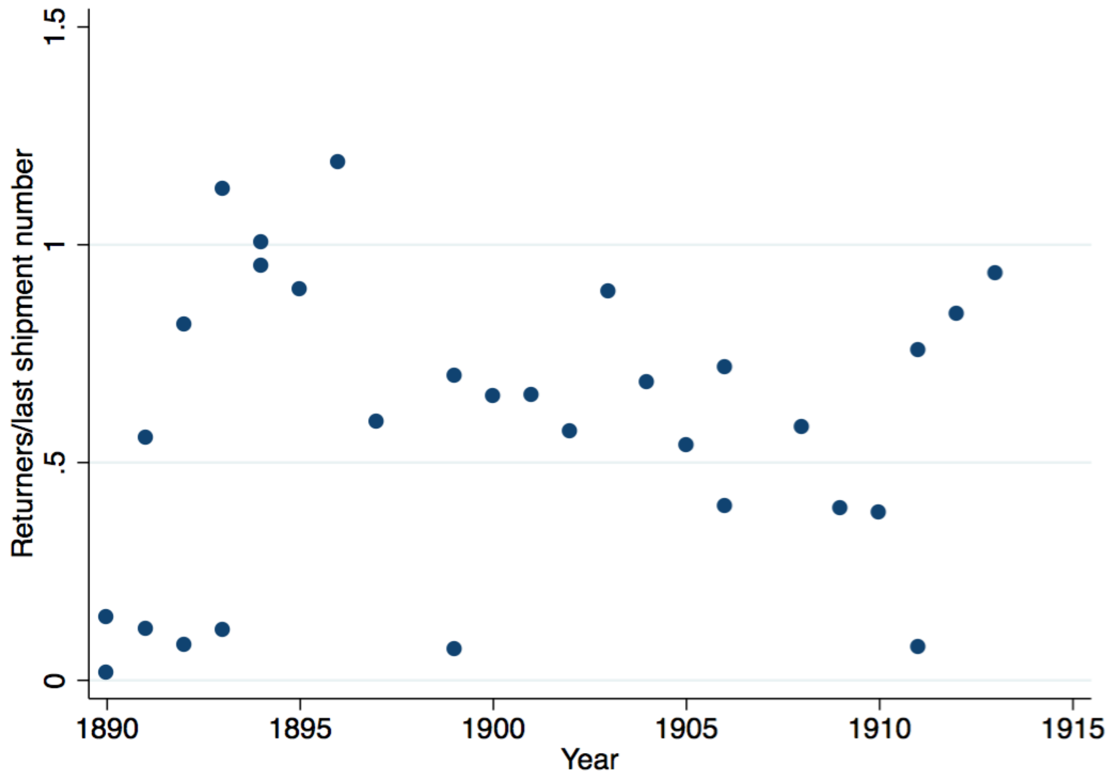


Table II.1: North Indian demographics, entrants through 1908

| | Mean | SD | N |
|---|---|---|---|
| Age (years) | 22.7 | (3.97) | 26594 |
| Female | .29 | (.45) | 26597 |
| United Provinces | .75 | (.43) | 26627 |
| Height (male, inches) | 64.17 | (2.37) | 18695 |
| Height (female, inches) | 59.21 | (2.27) | 7637 |
| *Castes* | | | |
| Brahman | .02 | (.15) | 26627 |
| Kshatriya | .15 | (.36) | 26627 |
| Ahir | .1 | (.3) | 26627 |
| Chamar | .14 | (.35) | 26627 |
| Muslim | .14 | (.35) | 26627 |

Table II.2: Return migration and mortality summary

| | Mean | SD | N |
|---|---|---|---|
| *Repatriated at the first opportunity conditional on survival* | | | |
| All | .03 | (.16) | 22364 |
| Male adult | .03 | (.18) | 15952 |
| Female adult | .01 | (.11) | 6412 |
| *Castes* | | | |
| Brahman | .06 | (.25) | 542 |
| Kshatriya | .04 | (.2) | 3136 |
| Ahir | .02 | (.15) | 2185 |
| Chamar | .01 | (.11) | 3136 |
| Muslim | .03 | (.16) | 3289 |
| *Death in years 0-5* | | | |
| All | .08 | (.27) | 26627 |
| Male adult | .07 | (.26) | 18869 |
| Female adult | .11 | (.31) | 7728 |
| *Castes* | | | |
| Brahman | .11 | (.31) | 648 |
| Kshatriya | .09 | (.29) | 3955 |
| Ahir | .09 | (.28) | 2592 |
| Chamar | .09 | (.28) | 3689 |
| Muslim | .06 | (.24) | 3811 |
| *Involuntary repatriation years 0-5* | | | |
| All | .01 | (.07) | 26627 |
| Male adult | 0 | (.07) | 18869 |
| Female adult | .01 | (.07) | 7728 |
| *Castes* | | | |
| Brahman | .01 | (.1) | 648 |
| Kshatriya | .01 | (.09) | 3955 |
| Ahir | .01 | (.08) | 2592 |
| Chamar | 0 | (.06) | 3689 |
| Muslim | 0 | (.07) | 3811 |

Table II.3: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | 0.00069 | 0.00069 | 0.00069 | 0.00069 |
| | (0.0016) | (0.0015) | (0.0015) | (0.0015) |
| Age squared | 0.0000042 | 0.0000064 | 0.0000052 | 0.0000074 |
| | (0.000031) | (0.000030) | (0.000031) | (0.000030) |
| Female | -0.017*** | -0.015*** | -0.0076*** | -0.0050** |
| | (0.0029) | (0.0026) | (0.0026) | (0.0024) |
| Passage price back (£) | -0.0048*** | -0.0048*** | -0.0048*** | -0.0048*** |
| | (0.00074) | (0.00072) | (0.00074) | (0.00072) |
| Brahman | 0.050*** | 0.049*** | 0.067*** | 0.067*** |
| | (0.011) | (0.011) | (0.016) | (0.016) |
| Kshatriya | 0.029*** | 0.028*** | 0.034*** | 0.034*** |
| | (0.0047) | (0.0049) | (0.0056) | (0.0059) |
| Ahir | 0.010** | 0.0093* | 0.013* | 0.012* |
| | (0.0050) | (0.0048) | (0.0071) | (0.0069) |
| Muslim | 0.0092*** | 0.0085*** | 0.0089*** | 0.0083** |
| | (0.0028) | (0.0029) | (0.0033) | (0.0035) |
| Non Chamar, other caste | 0.014*** | 0.013*** | 0.018*** | 0.017*** |
| | (0.0020) | (0.0022) | (0.0025) | (0.0026) |
| Married, spouse left behind | | 0.017** | | 0.017** |
| | | (0.0075) | | (0.0076) |
| Married, spouse came | | -0.0081* | | -0.0081* |
| | | (0.0044) | | (0.0045) |
| Brahman*female | | | -0.050*** | -0.051*** |
| | | | (0.019) | (0.019) |
| Kshatriya*female | | | -0.021*** | -0.021*** |
| | | | (0.0068) | (0.0068) |
| Ahir*female | | | -0.0072 | -0.0071 |
| | | | (0.0099) | (0.010) |
| Muslim*female | | | 0.0019 | 0.0015 |
| | | | (0.0057) | (0.0057) |
| Non Chamar, other caste*female | | | -0.012*** | -0.012*** |
| | | | (0.0035) | (0.0035) |
| Brahman | 10.34 | 10.14 | 14.07 | 13.9 |
| | (2.1) | (2.11) | (2.93) | (2.9) |
| Kshatriya | 5.92 | 5.78 | 7.2 | 7.08 |
| | (1.35) | (1.36) | (1.68) | (1.69) |
| Ahir | 2.16 | 1.92 | 2.74 | 2.5 |
| | (1.05) | (.96) | (1.49) | (1.39) |
| Muslim | 1.91 | 1.75 | 1.86 | 1.73 |
| | (.55) | (.58) | (.65) | (.69) |
| Non Chamar, other caste | 2.87 | 2.72 | 3.72 | 3.55 |
| | (.52) | (.52) | (.70) | (.68) |
| Observations | 22364 | 22364 | 22364 | 22364 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.4: Mortality in the first five years

| | (1) |
|---|---|
| Female | 0.31*** |
| | (0.092) |
| Age (years) | 0.11*** |
| | (0.031) |
| Age squared | -0.0017*** |
| | (0.00058) |
| Height (in) | -0.18 |
| | (0.14) |
| Height (in), squared | 0.0013 |
| | (0.0011) |
| Brahman | 0.15 |
| | (0.17) |
| Kshatriya | 0.017 |
| | (0.10) |
| Ahir | 0.038 |
| | (0.12) |
| Muslim | -0.33** |
| | (0.14) |
| Non Chamar, other caste | -0.13 |
| | (0.091) |
| Brahman*female | 0.15 |
| | (0.22) |
| Kshatriya*female | 0.095 |
| | (0.14) |
| Ahir*female | -0.032 |
| | (0.21) |
| Muslim*female | -0.012 |
| | (0.20) |
| Non Chamar, other caste*female | 0.085 |
| | (0.12) |
| Observations | 25593 |

Cox hazard model. SEs clustered by district in parentheses.
District FEs included. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table II.5: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | -0.0087** | -0.0087** | -0.0088** | -0.0088** |
| | (0.0040) | (0.0040) | (0.0040) | (0.0040) |
| Age squared | 0.00020** | 0.00021** | 0.00021** | 0.00021** |
| | (0.000083) | (0.000084) | (0.000083) | (0.000084) |
| Female | -0.019*** | -0.017*** | -0.012*** | -0.0096*** |
| | (0.0031) | (0.0028) | (0.0034) | (0.0033) |
| Passage price back (£) | -0.0027** | -0.0027** | -0.0027** | -0.0026** |
| | (0.0012) | (0.0012) | (0.0012) | (0.0012) |
| Brahman | 0.030* | 0.028* | 0.038** | 0.037* |
| | (0.016) | (0.016) | (0.019) | (0.019) |
| Kshatriya | 0.031*** | 0.030*** | 0.038*** | 0.037*** |
| | (0.0074) | (0.0076) | (0.0089) | (0.0091) |
| Ahir | 0.0024 | 0.00099 | 0.0016 | 0.00018 |
| | (0.0041) | (0.0043) | (0.0049) | (0.0050) |
| Muslim | 0.0036 | 0.0027 | 0.0036 | 0.0027 |
| | (0.0054) | (0.0054) | (0.0060) | (0.0061) |
| Non Chamar, other caste | 0.0090** | 0.0080** | 0.013*** | 0.012*** |
| | (0.0035) | (0.0035) | (0.0039) | (0.0039) |
| Married, spouse left behind | | 0.0090 | | 0.0093 |
| | | (0.013) | | (0.013) |
| Married, spouse came | | -0.012** | | -0.012** |
| | | (0.0049) | | (0.0049) |
| Brahman*female | | | -0.024 | -0.025 |
| | | | (0.020) | (0.020) |
| Kshatriya*female | | | -0.023** | -0.023** |
| | | | (0.0092) | (0.0094) |
| Ahir*female | | | 0.0052 | 0.0052 |
| | | | (0.0080) | (0.0080) |
| Muslim*female | | | 0.00030 | 0.00038 |
| | | | (0.0081) | (0.0081) |
| Non Chamar, other caste*female | | | -0.011*** | -0.011*** |
| | | | (0.0040) | (0.0039) |
| Brahman | 10.9 | 10.58 | 14.17 | 14.09 |
| | (6.4) | (6.54) | (7.97) | (8.17) |
| Kshatriya | 11.56 | 11.33 | 14.03 | 13.88 |
| | (5) | (5.01) | (6.23) | (6.27) |
| Ahir | .88 | .37 | .59 | .07 |
| | (1.59) | (1.64) | (1.85) | (1.9) |
| Muslim | 1.32 | 1.01 | 1.35 | 1.04 |
| | (1.89) | (1.96) | (2.16) | (2.25) |
| Non Chamar, other caste | 3.29 | 3.01 | 4.63 | 4.36 |
| | (1.49) | (1.44) | (2.01) | (1.92) |
| Observations | 10111 | 10111 | 10111 | 10111 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.6: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | -0.0057** | -0.0060** | -0.0057** | -0.0060** |
| | (0.0028) | (0.0027) | (0.0028) | (0.0027) |
| Age squared | 0.00015** | 0.00015** | 0.00015** | 0.00015** |
| | (0.000061) | (0.000060) | (0.000061) | (0.000061) |
| Female | -0.017*** | -0.015*** | -0.0075*** | -0.0049* |
| | (0.0025) | (0.0022) | (0.0025) | (0.0027) |
| Passage price back (£) | -0.0070*** | -0.0070*** | -0.0069*** | -0.0070*** |
| | (0.0015) | (0.0015) | (0.0015) | (0.0015) |
| Brahman | 0.055*** | 0.054*** | 0.076*** | 0.076*** |
| | (0.012) | (0.012) | (0.019) | (0.019) |
| Kshatriya | 0.032*** | 0.032*** | 0.038*** | 0.038*** |
| | (0.0053) | (0.0056) | (0.0064) | (0.0067) |
| Ahir | 0.013*** | 0.012*** | 0.015** | 0.014** |
| | (0.0047) | (0.0045) | (0.0066) | (0.0063) |
| Muslim | 0.014*** | 0.014*** | 0.014*** | 0.014*** |
| | (0.0042) | (0.0044) | (0.0045) | (0.0047) |
| Non Chamar, other caste | 0.017*** | 0.017*** | 0.021*** | 0.020*** |
| | (0.0031) | (0.0033) | (0.0035) | (0.0037) |
| Caste mortality prior year | 1.47** | 1.56*** | 1.48** | 1.56*** |
| | (0.57) | (0.57) | (0.57) | (0.57) |
| Married, spouse left behind | | 0.022*** | | 0.022*** |
| | | (0.0077) | | (0.0077) |
| Married, spouse came | | -0.0055 | | -0.0056 |
| | | (0.0050) | | (0.0051) |
| Brahman*female | | | -0.062*** | -0.063*** |
| | | | (0.024) | (0.024) |
| Kshatriya*female | | | -0.021*** | -0.021*** |
| | | | (0.0074) | (0.0075) |
| Ahir*female | | | -0.0051 | -0.0051 |
| | | | (0.0099) | (0.010) |
| Muslim*female | | | 0.0013 | 0.0010 |
| | | | (0.0063) | (0.0064) |
| Non Chamar, other caste*female | | | -0.012*** | -0.012*** |
| | | | (0.0038) | (0.0039) |
| Brahman | 7.81 | 7.72 | 10.94 | 10.86 |
| | (2.4) | (2.33) | (3.47) | (3.37) |
| Kshatriya | 4.52 | 4.5 | 5.41 | 5.4 |
| | (1.12) | (1.13) | (1.4) | (1.4) |
| Ahir | 1.84 | 1.75 | 2.16 | 2.06 |
| | (.89) | (.82) | (1.19) | (1.12) |
| Muslim | 2.04 | 2.03 | 2.03 | 2.03 |
| | (.45) | (.46) | (.55) | (.55) |
| Non Chamar, other caste | 2.42 | 2.38 | 2.98 | 2.94 |
| | (.44) | (.44) | (.6) | (.59) |
| Observations | 19171 | 19171 | 19171 | 19171 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.7: LPM: paid repatriation, repeated choices

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age | 0.0024*** | 0.0020*** | 0.0024*** | 0.0020*** |
| | (0.00047) | (0.00048) | (0.00047) | (0.00048) |
| Age squared | -0.000049*** | -0.000044*** | -0.000049*** | -0.000044*** |
| | (0.0000069) | (0.0000070) | (0.0000069) | (0.0000070) |
| female | -0.0093*** | -0.010*** | -0.0070*** | -0.0078*** |
| | (0.00078) | (0.00081) | (0.0014) | (0.0013) |
| Passage price back (£) | -0.0032*** | -0.0032*** | -0.0032*** | -0.0032*** |
| | (0.00069) | (0.00067) | (0.00069) | (0.00067) |
| Brahman | 0.016*** | 0.017*** | 0.019*** | 0.020*** |
| | (0.0031) | (0.0031) | (0.0037) | (0.0036) |
| Kshatriya | 0.015*** | 0.016*** | 0.017*** | 0.018*** |
| | (0.0020) | (0.0020) | (0.0022) | (0.0022) |
| Ahir | 0.0059** | 0.0068** | 0.0071** | 0.0080*** |
| | (0.0027) | (0.0026) | (0.0029) | (0.0028) |
| Muslim | 0.012*** | 0.013*** | 0.012*** | 0.013*** |
| | (0.0021) | (0.0020) | (0.0027) | (0.0026) |
| Non Chamar, other caste | 0.0081*** | 0.0087*** | 0.0088*** | 0.0094*** |
| | (0.0011) | (0.0011) | (0.0013) | (0.0013) |
| Married, spouse left behind | | 0.013*** | | 0.014*** |
| | | (0.0020) | | (0.0020) |
| Married, spouse came | | 0.0086*** | | 0.0086*** |
| | | (0.0014) | | (0.0014) |
| Brahman*female | | | -0.0078** | -0.0073* |
| | | | (0.0039) | (0.0040) |
| Kshatriya*female | | | -0.0061** | -0.0064** |
| | | | (0.0028) | (0.0027) |
| Ahir*female | | | -0.0041 | -0.0040 |
| | | | (0.0030) | (0.0029) |
| Muslim*female | | | -0.000015 | 0.000034 |
| | | | (0.0032) | (0.0032) |
| Non Chamar, other caste*female | | | -0.0019 | -0.0020 |
| | | | (0.0018) | (0.0017) |
| Constant | 0.0062 | 0.012 | 0.0051 | 0.011 |
| | (0.0082) | (0.0084) | (0.0084) | (0.0085) |
| Brahman | 4.91 | 5.42 | 5.83 | 6.27 |
| | (1.66) | (1.67) | (1.91) | (1.91) |
| Kshatriya | 4.76 | 5.05 | 5.31 | 5.63 |
| | (1.35) | (1.36) | (1.5) | (1.5) |
| Ahir | 1.82 | 2.1 | 2.21 | 2.49 |
| | (1.05) | (1.05) | (1.2) | (1.18) |
| Muslim | 3.72 | 3.95 | 3.74 | 3.96 |
| | (1.27) | (1.25) | (1.41) | (1.38) |
| Non Chamar, other caste | 2.52 | 2.72 | 2.73 | 2.93 |
| | (.73) | (.74) | (.85) | (.85) |
| Observations | 171439 | 171439 | 171439 | 171439 |

Dependent variable: paid repatriation ( price). District FEs included.

SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.



* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table II.8: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | 0.0010 | 0.0011 | 0.0011 | 0.0011 |
| | (0.0015) | (0.0014) | (0.0014) | (0.0014) |
| Age squared | -0.0000026 | -0.00000055 | -0.0000029 | -0.00000079 |
| | (0.000028) | (0.000028) | (0.000027) | (0.000027) |
| Female | -0.013*** | -0.011** | -0.0027 | -0.000027 |
| | (0.0047) | (0.0043) | (0.0036) | (0.0034) |
| Passage price back (£) | -0.0049*** | -0.0049*** | -0.0048*** | -0.0049*** |
| | (0.00077) | (0.00075) | (0.00077) | (0.00075) |
| Height (in) | 0.0091 | 0.0097 | 0.010 | 0.011 |
| | (0.0076) | (0.0075) | (0.0076) | (0.0076) |
| Height squared | -0.000067 | -0.000072 | -0.000076 | -0.000082 |
| | (0.000062) | (0.000062) | (0.000063) | (0.000062) |
| Brahman | 0.051*** | 0.050*** | 0.068*** | 0.068*** |
| | (0.011) | (0.011) | (0.016) | (0.016) |
| Kshatriya | 0.029*** | 0.029*** | 0.036*** | 0.035*** |
| | (0.0044) | (0.0046) | (0.0052) | (0.0054) |
| Ahir | 0.011** | 0.010* | 0.014* | 0.013* |
| | (0.0055) | (0.0052) | (0.0080) | (0.0078) |
| Muslim | 0.0098*** | 0.0091*** | 0.010*** | 0.0095*** |
| | (0.0027) | (0.0028) | (0.0033) | (0.0034) |
| Non Chamar, other caste | 0.015*** | 0.014*** | 0.019*** | 0.018*** |
| | (0.0019) | (0.0020) | (0.0026) | (0.0025) |
| Married, spouse left behind | | 0.017** | | 0.017** |
| | | (0.0074) | | (0.0075) |
| Married, spouse came | | -0.0081* | | -0.0081* |
| | | (0.0044) | | (0.0044) |
| Brahman*female | | | -0.051*** | -0.052*** |
| | | | (0.019) | (0.019) |
| Kshatriya*female | | | -0.022*** | -0.022*** |
| | | | (0.0067) | (0.0067) |
| Ahir*female | | | -0.0081 | -0.0080 |
| | | | (0.011) | (0.011) |
| Muslim*female | | | 0.00019 | -0.00018 |
| | | | (0.0059) | (0.0058) |
| Non Chamar, other caste*female | | | -0.013*** | -0.013*** |
| | | | (0.0040) | (0.0040) |
| Brahman | 10.41 | 10.2 | 14.12 | 13.94 |
| | (2.11) | (2.1) | (2.95) | (2.89) |
| Kshatriya | 6.01 | 5.87 | 7.38 | 7.25 |
| | (1.33) | (1.32) | (1.65) | (1.63) |
| Ahir | 2.31 | 2.07 | 2.95 | 2.71 |
| | (1.13) | (1.03) | (1.65) | (1.54) |
| Muslim | 2.02 | 1.86 | 2.09 | 1.95 |
| | (.52) | (.53) | (.64) | (.64) |
| Non Chamar, other caste | 3.04 | 2.88 | 3.95 | 3.78 |
| | (.52) | (.49) | (.73) | (.66) |
| Observations | 22120 | 22120 | 22120 | 22120 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Figure II.A1: Mortality prior 5 years by caste

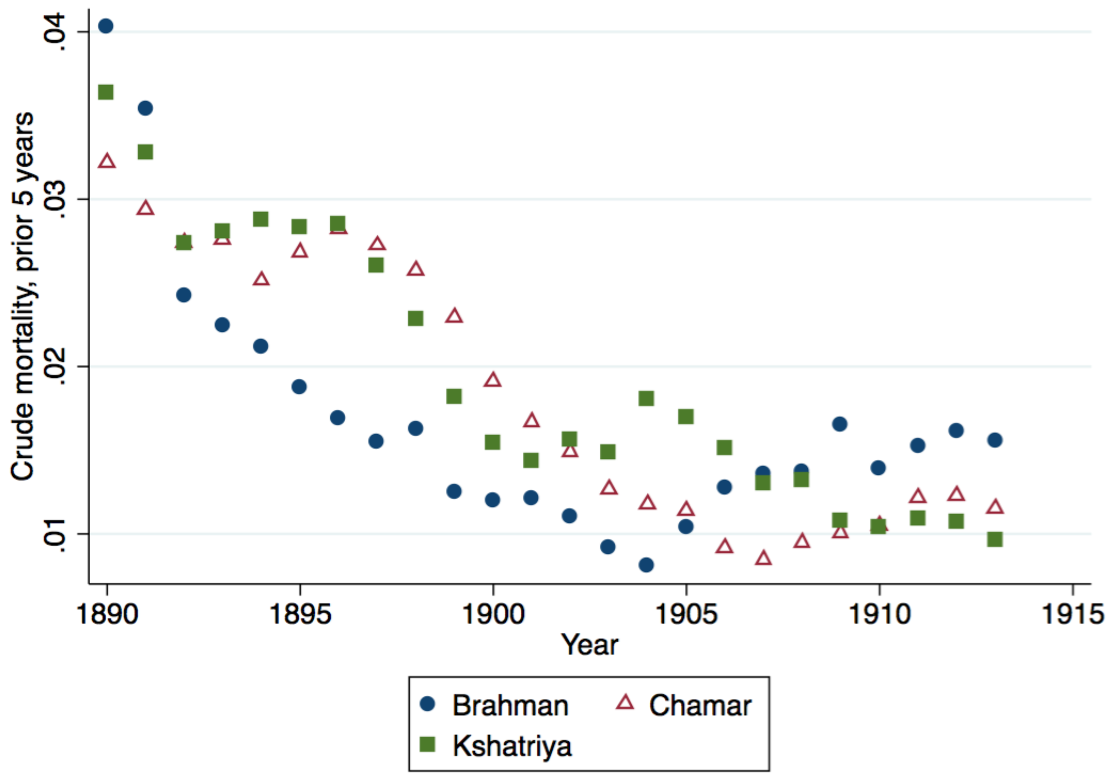


Table II.A1: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | -0.0032 | -0.0034 | -0.0032 | -0.0035 |
| | (0.0029) | (0.0028) | (0.0029) | (0.0028) |
| Age squared | 0.000092 | 0.000099 | 0.000094 | 0.00010 |
| | (0.000063) | (0.000062) | (0.000064) | (0.000063) |
| Female | -0.019*** | -0.017*** | -0.0066** | -0.0039 |
| | (0.0032) | (0.0028) | (0.0028) | (0.0028) |
| Passage price back (£) | -0.0051*** | -0.0051*** | -0.0051*** | -0.0051*** |
| | (0.00098) | (0.0010) | (0.00097) | (0.0010) |
| Brahman | 0.059*** | 0.059*** | 0.081*** | 0.081*** |
| | (0.012) | (0.013) | (0.017) | (0.017) |
| Kshatriya | 0.029*** | 0.029*** | 0.036*** | 0.036*** |
| | (0.0053) | (0.0057) | (0.0064) | (0.0067) |
| Ahir | 0.010* | 0.0095* | 0.014* | 0.013* |
| | (0.0054) | (0.0051) | (0.0078) | (0.0075) |
| Muslim | 0.0080** | 0.0075** | 0.0079* | 0.0075* |
| | (0.0035) | (0.0037) | (0.0040) | (0.0043) |
| Non Chamar, other caste | 0.015*** | 0.014*** | 0.020*** | 0.020*** |
| | (0.0024) | (0.0026) | (0.0030) | (0.0032) |
| Married, spouse left behind | | 0.019** | | 0.019** |
| | | (0.0087) | | (0.0088) |
| Married, spouse came | | -0.0071 | | -0.0072 |
| | | (0.0054) | | (0.0055) |
| Brahman*female | | | -0.063*** | -0.064*** |
| | | | (0.021) | (0.021) |
| Kshatriya*female | | | -0.025*** | -0.025*** |
| | | | (0.0073) | (0.0074) |
| Ahir*female | | | -0.0083 | -0.0083 |
| | | | (0.012) | (0.012) |
| Muslim*female | | | 0.0014 | 0.0011 |
| | | | (0.0067) | (0.0067) |
| Non Chamar, other caste*female | | | -0.016*** | -0.016*** |
| | | | (0.0044) | (0.0045) |
| Brahman | 11.64 | 11.49 | 15.98 | 15.87 |
| | (2.95) | (3.1) | (4.16) | (4.28) |
| Kshatriya | 5.74 | 5.65 | 7.18 | 7.11 |
| | (1.58) | (1.7) | (1.98) | (2.12) |
| Ahir | 2.03 | 1.85 | 2.67 | 2.5 |
| | (1.03) | (.96) | (1.51) | (1.44) |
| Muslim | 1.58 | 1.47 | 1.55 | 1.47 |
| | (.82) | (.88) | (.9) | (.96) |
| Non Chamar, other caste | 2.93 | 2.81 | 3.97 | 3.85 |
| | (.66) | (.73) | (.84) | (.9) |
| Observations | 19269 | 19269 | 19269 | 19269 |

Dependent variable: paid repatriation (strict price) at five years. District FEs included.

SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.A2: LPM: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | -0.0061** | -0.0063** | -0.0061** | -0.0063** |
| | (0.0028) | (0.0027) | (0.0028) | (0.0028) |
| Age squared | 0.00015** | 0.00016*** | 0.00015** | 0.00016** |
| | (0.000061) | (0.000061) | (0.000062) | (0.000062) |
| Female | -0.017*** | -0.015*** | -0.0073*** | -0.0048* |
| | (0.0025) | (0.0022) | (0.0025) | (0.0027) |
| Passage price back (£) | -0.0059*** | -0.0060*** | -0.0058*** | -0.0059*** |
| | (0.0012) | (0.0012) | (0.0012) | (0.0012) |
| Brahman | 0.059*** | 0.059*** | 0.080*** | 0.080*** |
| | (0.013) | (0.013) | (0.019) | (0.019) |
| Kshatriya | 0.029*** | 0.029*** | 0.035*** | 0.035*** |
| | (0.0053) | (0.0056) | (0.0065) | (0.0067) |
| Ahir | 0.012** | 0.011** | 0.014** | 0.013** |
| | (0.0048) | (0.0046) | (0.0067) | (0.0064) |
| Muslim | 0.010*** | 0.011*** | 0.010** | 0.011** |
| | (0.0036) | (0.0037) | (0.0041) | (0.0042) |
| Non Chamar, other caste | 0.015*** | 0.015*** | 0.019*** | 0.019*** |
| | (0.0028) | (0.0030) | (0.0032) | (0.0035) |
| Caste mortality past 5 years | 0.72* | 0.87** | 0.71* | 0.86** |
| | (0.40) | (0.40) | (0.40) | (0.40) |
| Married, spouse left behind | | 0.022*** | | 0.022*** |
| | | (0.0077) | | (0.0078) |
| Married, spouse came | | -0.0052 | | -0.0053 |
| | | (0.0049) | | (0.0050) |
| Brahman*female | | | -0.062*** | -0.062*** |
| | | | (0.023) | (0.023) |
| Kshatriya*female | | | -0.021*** | -0.021*** |
| | | | (0.0075) | (0.0075) |
| Ahir*female | | | -0.0056 | -0.0056 |
| | | | (0.0098) | (0.0099) |
| Muslim*female | | | 0.00085 | 0.00060 |
| | | | (0.0064) | (0.0064) |
| Non Chamar, other caste*female | | | -0.012*** | -0.012*** |
| | | | (0.0038) | (0.0039) |
| Brahman | 10 | 9.83 | 13.76 | 13.56 |
| | (2.73) | (2.62) | (3.99) | (3.84) |
| Kshatriya | 5 | 4.9 | 6.08 | 5.98 |
| | (1.35) | (1.33) | (1.69) | (1.66) |
| Ahir | 2.01 | 1.87 | 2.42 | 2.27 |
| | (1.03) | (.95) | (1.39) | (1.29) |
| Muslim | 1.76 | 1.78 | 1.76 | 1.8 |
| | (.53) | (.53) | (.65) | (.65) |
| Non Chamar, other caste | 2.52 | 2.48 | 3.19 | 3.14 |
| | (.51) | (.51) | (.70) | (.69) |
| Observations | 19171 | 19171 | 19171 | 19171 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.A3: Logit: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | 0.060 | 0.061 | 0.059 | 0.059 |
| | (0.059) | (0.052) | (0.060) | (0.053) |
| Age squared | -0.00065 | -0.00057 | -0.00060 | -0.00051 |
| | (0.0011) | (0.00099) | (0.0011) | (0.0010) |
| Female | -0.87*** | -0.77*** | -0.79*** | -0.67** |
| | (0.15) | (0.13) | (0.26) | (0.26) |
| Passage price back (£) | -0.26*** | -0.26*** | -0.25*** | -0.26*** |
| | (0.041) | (0.039) | (0.041) | (0.040) |
| Brahman | 1.42*** | 1.39*** | 1.51*** | 1.49*** |
| | (0.21) | (0.23) | (0.24) | (0.24) |
| Kshatriya | 1.16*** | 1.14*** | 1.18*** | 1.16*** |
| | (0.19) | (0.20) | (0.18) | (0.20) |
| Ahir | 0.57*** | 0.52** | 0.57** | 0.53** |
| | (0.21) | (0.21) | (0.26) | (0.26) |
| Muslim | 0.52*** | 0.50*** | 0.42*** | 0.40** |
| | (0.17) | (0.18) | (0.16) | (0.17) |
| Non Chamar, other caste | 0.68*** | 0.65*** | 0.72*** | 0.69*** |
| | (0.14) | (0.15) | (0.13) | (0.14) |
| Married, spouse left behind | | 0.52*** | | 0.52*** |
| | | (0.18) | | (0.18) |
| Married, spouse came | | -0.36 | | -0.36 |
| | | (0.23) | | (0.23) |
| Brahman*female | | | -0.48 | -0.53 |
| | | | (0.46) | (0.46) |
| Kshatriya*female | | | -0.11 | -0.12 |
| | | | (0.41) | (0.41) |
| Ahir*female | | | 0.0090 | 0.0092 |
| | | | (0.66) | (0.67) |
| Muslim*female | | | 0.48 | 0.46 |
| | | | (0.43) | (0.42) |
| Non Chamar, other caste*female | | | -0.29 | -0.29 |
| | | | (0.34) | (0.34) |
| Brahman | 5.57 | 5.35 | 5.92 | 5.74 |
| | (.89) | (.87) | (.97) | (.93) |
| Kshatriya | 4.55 | 4.39 | 4.62 | 4.47 |
| | (.89) | (.88) | (.94) | (.92) |
| Ahir | 2.22 | 2 | 2.24 | 2.03 |
| | (.84) | (.77) | (1.06) | (.98) |
| Muslim | 2.04 | 1.92 | 1.65 | 1.55 |
| | (.58) | (.62) | (.56) | (.59) |
| Non Chamar, other caste | 2.66 | 2.51 | 2.83 | 2.67 |
| | (.55) | (.56) | (.58) | (.56) |
| Observations | 23380 | 23380 | 23380 | 23380 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included. SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table II.A4: Probit: paid repatriation at first opportunity

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Age (years) | 0.021 | 0.020 | 0.020 | 0.020 |
| | (0.027) | (0.024) | (0.027) | (0.024) |
| Age squared | -0.00018 | -0.00013 | -0.00016 | -0.00011 |
| | (0.00052) | (0.00048) | (0.00052) | (0.00048) |
| Female | -0.35*** | -0.30*** | -0.31*** | -0.25** |
| | (0.058) | (0.054) | (0.098) | (0.099) |
| Passage price back (£) | -0.10*** | -0.11*** | -0.10*** | -0.10*** |
| | (0.016) | (0.015) | (0.016) | (0.016) |
| Brahman | 0.61*** | 0.60*** | 0.67*** | 0.67*** |
| | (0.095) | (0.10) | (0.11) | (0.11) |
| Kshatriya | 0.49*** | 0.49*** | 0.50*** | 0.50*** |
| | (0.079) | (0.085) | (0.079) | (0.085) |
| Ahir | 0.24*** | 0.22*** | 0.24** | 0.22** |
| | (0.085) | (0.082) | (0.11) | (0.10) |
| Muslim | 0.21*** | 0.20*** | 0.17** | 0.16** |
| | (0.069) | (0.074) | (0.065) | (0.069) |
| Non Chamar, other caste | 0.28*** | 0.27*** | 0.30*** | 0.29*** |
| | (0.055) | (0.060) | (0.052) | (0.056) |
| Married, spouse left behind | | 0.24*** | | 0.24*** |
| | | (0.083) | | (0.084) |
| Married, spouse came | | -0.15* | | -0.15* |
| | | (0.091) | | (0.092) |
| Brahman*female | | | -0.26 | -0.29 |
| | | | (0.19) | (0.19) |
| Kshatriya*female | | | -0.087 | -0.094 |
| | | | (0.16) | (0.16) |
| Ahir*female | | | 0.019 | 0.011 |
| | | | (0.26) | (0.26) |
| Muslim*female | | | 0.19 | 0.18 |
| | | | (0.17) | (0.17) |
| Non Chamar, other caste*female | | | -0.13 | -0.13 |
| | | | (0.13) | (0.13) |
| Brahman | 5.92 | 5.71 | 6.5 | 6.35 |
| | (.97) | (.97) | (1.11) | (1.08) |
| Kshatriya | 4.75 | 4.62 | 4.9 | 4.78 |
| | (.91) | (.93) | (.99) | (.99) |
| Ahir | 2.31 | 2.12 | 2.32 | 2.14 |
| | (.84) | (.77) | (1.09) | (.99) |
| Muslim | 2.06 | 1.94 | 1.6 | 1.51 |
| | (.57) | (.62) | (.56) | (.59) |
| Non Chamar, other caste | 2.7 | 2.54 | 2.91 | 2.76 |
| | (.53) | (.55) | (.58) | (.57) |
| Observations | 23380 | 23380 | 23380 | 23380 |

Dependent variable: paid repatriation (flexible price) at five years. District FEs included.
SEs for coefficients clustered by district. SEs for caste values calculated via the delta method.
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

CHAPTER 3

# Foreign/native wage differentials and return migration in Germany[1]

## 1 Introduction

International migration is a salient issue in modern-day economies, as millions of workers travel across national borders to work in other countries. However, the choice to migrate is often endogenous, and it is hard to determine who migrates and why (e.g., Borjas [1987]). The latent 'intention to migrate' does not always turn into migration. The benefit of migrating may be over- or understated depending on the features of the sending and recipient labor markets.

Despite this caveat, the labor market outcomes of migrant workers are of interest. Depending on the institutional regime and composition of the origin and recipient workforces, multiple outcomes are possible, such as segmentation into a two-tier labor market (one native, one migrant) and divergence in wages. Differentials in wages may result from different causes. Differences in characteristics, especially human capital, may lead to different wages. For example, if foreign workers enter enter with lower education relative to the natives, they may earn less.

Discrimination may also play a role (see Becker [1957] for a discussion on discrimination more generally between a privileged in-group and an out-group). Some form of discrimination may occur to alter wages of one group. Becker's view implies that discrimination can occur in several dimensions, including price (wage) and labor-force segmentation.

[1]This study uses the weakly anonymous Biographical data of selected insurance agencies in Germany (BASiD) (Version 1951 - 2009). Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.

This study uses the factually anonymous Sample of Integrated Labour Market Biographies (version 1975 - 2010). Data access was provided via a Scientific Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Return migration affects both the impact of migrants and the differences between migrants and natives. Migrants may return to their home countries after hitting a certain income target, for example, or migrants unable to compete in the recipient labor market may choose to return home, which in turn biases estimates of the return to migration and of differentials between natives and foreign workers. The form of selection in permanent migration is important for policymakers in determining the long-run impact of migration. I focus on wage differentials and ask two related questions. First, is there an observed differential between foreigners and natives? Second, do differentials persist over time with those migrants who choose to stay?

I propose a method to measure the impact of return migration in Germany and to determine which part of the migrant worker distribution returns home. I examine wage differentials using two long-run administrative data sets from Germany, described below. Representative data stretching back to the 1970s in West Germany answer the first and show that wage differentials exist. Differences in human capital levels and remuneration lead to these differentials. To examine the second question, the nature of selection, I exploit differences in sampling rules between the two data sets to characterize the form of attrition and the remaining wage differentials. I find that permanent migrants are negatively selected and wage differentials grow over time.

Germany is an ideal location to study these issues. First, it is a large economy with a high number of immigrants and non-immigrant foreign workers. Per the Census of 2011, over 7 millions individuals (8.7% of the population) were foreign. In several states (*Lander*), over one in eight individuals are foreign.[10]

Second, Germany's modern immigration history stretches back for several decades, and so it is possible to observe foreign workers completing much of their working careers in Germany. West Germany's post-Second World War immigration boom stemmed from a chronic labor shortage and a desire to re-integrate West Germany with western Europe. Agreements with Italy (1955), Spain (1960), Greece (1960), Turkey (1961), Portugal (1964), and Yugoslavia (1968) in the 1950s and 1960s paved the way for mass immigration, usually of low- or unskilled workers, into West Germany through the *Gastarbeiterprogramm* (literally, "guest worker program"). Although the program was originally touted as a temporary solution to West Germany's post-war labor shortage, sustained immigration and settlement by immigrant groups in West Germany became the norm by the late 1970s/early 1980s.

Third, rich administrative data, described below, contain the requisite variables for measuring the impact of return migration. Germany collects data daily on workers for social security payments, and differing data sets can be exploited to understand migration topics. Previous research on wage differentials in Germany has pointed to the existence of differentials between foreigners and Germans. However, these studies have either not used administrative data or are now outdated. Changes in the economy, particularly after the adoption of the euro and new rules on migration, make an update necessary.

The evidence, while not all consistent, points to differentials between German[2] and foreign workers. However, early studies rely on small cross-sectional samples that cannot examine the role of return migration, and later samples do not focus on long-run trends. Furthermore, no study takes advantage of the long-run pension data.

Prior research using the German Socio-Economic Panel (GSOEP), a survey starting in the 1980s, find wage differentials between German and foreigners and a role for human capital. Pischke (1992) finds that foreign-born workers earn less than native-born Germans on the whole. However, he argues that this is consistent with a model of sorting by industry and that most foreign-born workers work in blue-collar industries. Within these industries, though, there are no wage differentials. Similarly, Schmidt (1997) finds that that immigrant wage differentials disappear once foreign-born workers are compared to similar native-born workers, i.e., controlling for education and occupation, foreign-born workers' earnings look similar to native-born workers' earnings. However, both rely on smaller surveys that may precisely suffer from attrition due to return migration.

Dustmann (1997a) also uses the GSOEP to explore female labor supply differences between native-born Germans and immigrants. He asks how immigrant attitudes towards return migration affect labor-market supply, and he builds a model in which immigrants incorporate this return plan. Consistent with the model, he finds that immigrants who plan to return participate at higher rates in in order to maximize their earnings abroad. However, the intent to return may differ from actual return migration.

Velling (1995) uses a 1% sample of all workers in 1989 to study male wage gaps. Using this large cross-section of workers, Velling finds that there is a wage differential between foreign-born and native-born workers. However, he argues that this can be explained by differences in human

[2]For the purposes of this paper, a native German is a native-born worker.

capital accumulation.

Pischke and Velling (1997) study local labor markets in West Germany and the impact of immigration. The unit of analysis is the county, and the authors examine how county-level macroeconomic indicators change with more immigrants. Although immigration choice is non-random and poses endogeneity issues, they argue that prior economic indicators are sufficient controls for immigrant location and sorting. They find no detrimental effect on native-born German workers and no evidence of out-migration of native-born Germans from high immigrant areas.[19]

Bartolucci (2009) takes a markedly different approach using administrative data from IAB. He uses matched employer-employee panel data (German acronym: LIAB) to track firms over time and their wages paid to different types of workers. He finds that immigrants are positively segregated into good firms but are subsequently paid 16% less than native Germans. This evidence does not support a taste-based discrimination model, although good firms discriminate less, but does support statistical discrimination: firms cannot (or do not) differentiate between foreign workers and pay the prevailing pooling wage.

D'Amuri, Ottaviano, and Peri (2010) use administrative employee data from the IAB. They focus on the years 1987-2001 because they argue that these are primary immigrant influx years. D'Amuri et al., like Pischke and Velling, find no evidence that increased immigration to West Germany/Germany in this period had little impact on native Germans' wages and employment but a large impact on prior immigrants. However, they only examine cross-sectional data rather than a richer panel set. I use the same data as D'Amuri et al. but I use all the available years, not just 1987-2001.

Several studies theorize how individuals make return migration decisions and to estimate the rate of return migration. Galor and Stark (1991) model return migration and argue that foreign workers with a positive probability of returning will exert higher effort (and thus earn more) than comparable native workers. Dustmann (1997b) complements Dustmann (1997a) and builds on Galor and Stark's model. He links return behavior to the risk of aggregate shocks; levels of uncertainty affect precautionary saving, consumption, and return behavior. He argues that migrants optimally choose time in the recipient country to maximize earnings there to re-invest or consume in the home market.

Merkle and Zimmermann (1992) examine remittances and other economic variables of interest amongst guest workers in Germany using the GSOEP. They find that return migration intentions significantly affect remittances; remittances fall as migrants choose to spend more time in Germany. This result is consistent with Galor and Stark's model.

Borjas and Bratsberg (1996) begin from the Roy Model and examine return migration decisions. They assert that the form of selection into migration amplifies subsequent return migration: if migrants are positively selected, then the less skilled ones return; the opposite holds for negative selection. However, this requires knowledge about the distribution of the sending area in order to understand migration motivations.

In another context, Yang (2006) examines exchange rate shocks for Filipino migrants in various countries around the world. He finds that the life-cycle theory of migration, in which migrants choose a destination and migrate there for their working lives due to higher wages, fits the data best.

I contribute to both strands of the literature. First, I use two long-run data sets of workers and so I can examine a longer time span. The data I use are administrative, not self-reported, and I can follow workers for many years rather than using points in time. Furthermore, I focus on local labor markets as the key geographic area of interest; this means that labor markets that cross over state boundaries, such as Berlin spilling over into Brandenburg, are correctly accounted. Second, due to sampling differences in the data sets I use, I can ask similar but different questions in order to reinforce my results from each. This provides an additional test of Galor and Stark: their model predicts that wage differentials should increase over time as some migrants return.

I approach the data agnostically and posit my results as a test of these various models. First, Galor and Stark (1991) imply that migrants who remain should be worse off than comparable native workers. Second, Borjas (1987) argues for selection in migration from one side of the distribution, and Borjas and Bratsberg (1996) argue that return migration amplifies this. My data enable me to test both.

## 2 Data

I utilize two large administrative data sets from the German Institute for Employment Research of the Federal Employment Agency (German acronym: IAB). Employers are required to report daily wages and other information to the Federal Employment Agency for pension and other purposes. The first data set, the Sample of Integrated Labour Market Biographies (German acronym: SIAB), is a 2% random sample of employees and benefits recipients in Germany from 1975-2010. The entire histories (hence, labor market biographies) of all sampled individuals are included in the data set to create a panel. Because each worker*day has an equal probability of being selected (the probability of being selected does not vary by amount of time worked), the data are representative of the employed population across all years.

The second data set is the Biographical data of selected insurance agencies in Germany (German acronym: BASiD). This is a panel data set consisting of a 1% random sample of all participants born in 1940 or later still in the German pension system as of 31 December 2007. The data stretch back until the 1950s. BASiD can be used to gather retrospective work-history data for workers, but historical analysis suffers from attrition: a person who did not survive in the pension system until 31 December 2007 is automatically excluded. Older workers have more periods at risk of death, and older migrants have the additional risk of returning home. The data become less representative further back in time, and the effect is mechanically amplified.

The key difference between SIAB and BASiD is the sample population: SIAB contains workers who left the labor force (e.g., migrants who returned home), whereas BASiD requires that the individuals still be in the pension system. By construction, BASiD is not representative as a panel or for any historical analyses, e.g., whether or not there actually were wage differentials in 1989; SIAB is instead used for that question. However, this then means that observed differences between the two data sets arise from sampling. SIAB may be interpreted as the 'true' representation of the German labor market and deviations seen in BASiD reflect the impact of return migration.

Key variables of interest include daily wages, citizenship/nationality, linked establishment (i.e., place of work) characteristics, and job type. I classify jobs using a 12-category system per Blossfeld (1987), which was designed for use with similar German data. I also use 105 local labor markets as defined by Kropp and Schwengler (2012) to capture the effects across different areas of West Germany/Germany. These are defined as of the late 2000s; I assume that they hold also going back in time to 1975. While some areas of local labor markets were clearly not included prior to unification in 1990 (e.g., West Berlin was isolated from its modern local labor market), this truncation should not create problems. The lack of East German data assists here. First, major labor markets in West Germany, such as in the Ruhr Valley, were unaffected geographically by unification. Second, given the spatial dimension of local labor markets, changes would have to entail different commuter patterns and a shift to an alternate labor market. Peripheral areas near the boundaries between local labor markets may have switched, but they are by definition small relative to the total size of the labor force.

For historical reasons, German administrative data does not contain information on ethnicity or place of birth but only nationality/citizenship. Therefore, there is a potential identification problem of naturalized foreign-born workers. In both datasets, though, there are often multiple records for a given worker. Some workers indicate both German citizenship and another citizenship. I create three categories of workers: German-only, foreign-only, and German-with-foreign. While there could be underreporting for the last category due to foreigners becoming German citizens without a simultaneous change in reported nationality, I argue that this provides a check between potentially more and less assimilated immigrants.

One caveat with respect to nationality relates to Germany's jus sanguinis laws; children born of foreigners may also be classified as foreign and should be in many cases. However, I address this by examining two groups against a German group: individuals who are listed only as foreigners, and individuals who are listed as foreigners and Germans at different points. This allows me to differentiate somewhat between different 'types' of foreigners in Germany in addition to nationality. Furthermore, it is not clear that misclassification creates problems. Children born to foreigners classified as German, i.e., children who would have been German citizens under jus soli, are in the correct group. Children classified as foreign or German*foreign should bias the differential towards *not* being significant. Therefore, the bias works in my favor, since any differential must overcome this problem.

I limit my sample to individuals working on 30 June of a given sample year. This date is chosen to correspond to the reporting date for establishments. Thus, I capture the correct, linked establishment data for the relevant population of workers. I calculate real wages using macroeconomic data for Germany provided by the IAB.

Descriptive statistics for the two data sets are given above. Table III.1 gives descriptive statistics at the individual level. I define max education as the maximum education reported by an individual across her histories; education variables are created according to the imputation method in Fitzenberger et al (2005). The omitted category is applied/technical university(*Fachhochschule*). This differs from university (*Universität*), and so this separation is maintained here. The two data sets appear generally similar in the demographic characteristics of the individuals analyzed. SIAB and BASiD have the same proportion of foreigners, but BASiD contains more German*foreign individuals and fewer only Germans. BASiD contains more females proportionally than SAIB. In particular, the BASiD sample contains a higher proportion of foreigners who are female than SIAB. I attribute this to different settlement patterns and return migration, since males are (in the literature) more likely to migrate temporarily

Table III.2 gives descriptive statistics at the job spell level. Within each data set, real wages for German-only individuals are higher than for foreign-only and German*foreign individuals, both overall and for males only. Real wage levels are slightly lower in BASiD than in SIAB, although workers tend to be older in SIAB than in BASiD. The difference in ages may result in the observed difference in real wages.

SIAB has a linked establishment variable on the composition of the labor force for the establishment. This is reported only for SIAB; unfortunately, this variable was not linked for BASiD in its creation. The differences are stark across the different groups: while German-only individuals tend to work with other Germans, foreign-only individuals work in more mixed firms (70% vs over 94% for German-only). Even the German*foreign individuals work in more mixed firms (77%). This fits in with Bartolucci's claim that workers segregate into firms, and it shows initial evidence of clustering by foreign workers into certain firms.

Local labor markets form an important secondary focus of this study. I incorporate local labor market information in order to reduce potential omitted variable bias problems, such as differential wages in an area and different compositions of the workforce/establishments. Both data sets contain location information, and I use consistent measures of labor markets across the two.

## 3 Empirical model: SIAB

### 3.1 Pooled regressions

Due to privacy concerns and data restrictions, I cannot merge the two main data sets together. Instead, I analyze them separately and compare the overall results. For SIAB, I first run a basic pooled Mincer-style regression:

$$\log(daily\_wage_{it}) = \beta_0 + \beta_1 I_{foreign;i,t} + \beta_2 I_{German*foreign;i,t} + \Gamma_1 Z_{i,t} + \Gamma_2 Year + \Gamma_3 Geo + \epsilon_{it} \quad (1)$$

where $Z$ is a vector containing person-level characteristics such as sex, age at spell, education, and job type; $Year$ is a vector of year fixed effects; and $Geo$ is a vector of geographic fixed effects (either state or local labor market). I run the specification multiple times with different mixes of variables in $Z$. I cluster standard errors by person.

The key coefficients of interest are $\beta_1$ and $\beta_2$. Recall that $\beta_1, \beta_2$ are not mutually exclusive. As discussed above, the nationality variable can take on several different values and so I define three groups: German only (omitted), foreign only, and German*foreign. $\beta_1$ measures the impact of being a foreigner and $\beta_2$ measures the additional difference due to being in the German*foreign category.

Table III.3 shows results with geographic fixed effects based on local labor markets. The omitted category is no degree, German, male workers. $\beta_2$ is negative and significant at the 1% level in all specifications. Thus, German*foreign workers a penalty for being foreign and German. However, $\beta_1$ is positive and significant at the 1% level in all specifications. Both $\beta_1$ and $\beta_2$, though, capture the mean differences for no education. This means that foreigners with no education receive a bonus for being foreign, i.e., they earn more than Germans conditional on the other covariates. At higher levels of education, the interaction terms on education with foreign and German*foreign are are negative: at each education level, foreign workers earn less than German workers.

The education variables mostly show the expected sign and magnitude. While wage does not increase monotonically in education, since workers with a high school degree and no training earn less than those without a degree, the overall relationship is of increasing wages with increasing education. In column 7, university graduates earn 27% more than those without a degree. Women earn less than men in all specifications, though the magnitude declines to 27% less in column 7 from almost 47% less in column 1. There are positive but decreasing returns to age, which proxies for experience. Interestingly, workers who work in firms with a higher proportion of native German workers earn more. This again supports Bartolucci's work on firm sorting since native German workers, who overwhelmingly work with other native Germans, reap the benefits. I attribute the slight increase in $\beta_1$ from columns 6 to 7 to the inclusion of the proportion native German.

The introduction of additional controls for local labor market, year, firm type, and worker type reduce the magnitudes of the covariates of interest and yield more precise estimates. The drop in absolute value of both $\beta_1$ and $\beta_2$ from columns 1 to 2 and columns 5 to 6, for instance, results from the inclusion of a worker category. This provides evidence for sorting by citizenship into different job types, with native Germans sorting into more remunerative occupations.

Column 7 is the preferred specification. Note here that $\beta_1 + \beta_2 \approx 0$. Thus, taking into account all of the controls, a native German worker with no degree earns similarly to a German*foreign worker with no degree. At high levels of education, the interaction term between foreign and the education variable can cancel out the foreign bonus from $\beta_1$. For example, a native German, university-educated worker earns only slightly more than a foreign, university-educated worker.

In order to explore the possibility of heterogeneous outcomes across national origin, I split the foreign variable into several parts: Turkey, Italy, Greece, Poland, former Yugoslavia, and all other. The abbreviated results are shown in Table III.4. For space, I omit reporting the interacted education variables. In large part, the disaggregated results here resemble the results in Table III.3: the coefficients on the various country variables (e.g., Turkey, Italy) are positive and significant at the 1% level while the interactions are negative and significant at the 1% level. One notable difference lies with Poland: the Poland variable changes sign and, in column 7, is negative and significant at the 1% level. The interaction is negative and significant at the 1% level in columns 1-5 but then close to zero and insignificant in columns 6-7. This could point to trend differences with Polish immigration, particularly given Poland's accession to the EU in the early 2000s. In column 7, the combined 'other' group has a coefficient (2.2%) less than the analogous overall foreign coefficient (5.6%) in Table III.3. However, this group is relatively small given that most foreign workers come from the specific countries included.

The specifications for both Tables III.3 and III.4 above are for all workers, i.e., regardless of the hours worked. I run robustness checks with workers in explicitly full-time work, and the results are substantively the same.[3]

Overall, the picture emerges of foreigner workers remunerated more highly at the lowest education level (no degree) than their native German counterparts, but otherwise similar or higher

[3] Workers in explicitly full-time work account for over 77% of worker-spell observations.

remuneration for native German workers. German*foreign workers look similar to native German workers at the lowest education and to foreign workers at higher education levels. The higher remuneration to foreign workers at low education levels could reflect mismeasurement of education, unobserved experience in the sending country, or a pooling equilibrium by employers due to uncertainty about the educational qualifications of foreign workers. Based on the the distribution of education from Table III.1, it appears that the higher remuneration to native German workers seen in Table III.2 comes mostly from the higher returns to education. I explore this further below.

### 3.2 Oaxaca-Blinder regressions

Oaxaca (1973) and Blinder (1973) emphasize that some wage differentials may be explainable, such as human capital differences, and others are not explainable. The latter are often interpreted as discrimination in a Becker sense: discrimination against a group leads to lower wages. While Oaxaca-Blinder regressions are important largely for their empirical application, the division between explained and unexplained differences provides key insight for interpreting differentials.

Thus, I next run Oaxaca-Blinder regressions using the same specifications as in columns 2-7 of Tables III.3. I utilize Jann's `oaxaca` command in Stata. These are linear regressions where the base group is German-only and the comparison group is the union of foreign-only and German*foreign. Thus, the comparison group differs from the main linear regressions because the two comparison groups above are pooled together. For space, the full output is omitted and only the grouped differences are reported.

Table III.5 provides the Oaxaca-Blinder regressions. Differences in coefficients appear to drive the result of a positive difference between German-only workers and foreign workers. Coefficients are positive and significant at the 1% level. This means that German-only workers receive a higher return to their characteristics than the any-foreign group. The magnitude of this difference increases in columns 4 and 6, the preferred column, which include proportion native German workers. Age, a proxy for experience, largely drives the differences rather than education; this contrasts somewhat with the regular regressions. Overall, the percent change in real wages is lower for a foreigner (or German*foreign) than for a German-only worker with identical observable characteristics.

Furthermore, the value of endowments is higher (significant at the 1% level) for the native German group in all specifications except columns 1 and 3; these two columns do not include worker type fixed effects. This shows that native Germans sort into higher remunerative job types than foreigners, and this seems to drive the endowment effect. When worker type is not included, the any-foreign group actually earns more. This is partly surprising, since the descriptive statistics in Table III.1 show lower levels of human capital, for example, for foreign and German*foreign workers, and Table III.5 showed that remuneration was higher for higher levels of education. This can be explained largely by the impact of local labor markets (not shown): any-foreign workers sort into higher-paying labor markets. Job type, though, is more important.

The evidence points to wage differentials between German-only workers and others. A more accurate measure of spatial correlation, local labor markets, shows this in Table **??**. The Oaxaca-Blinder regressions, while pooled for a slightly different sample than in the basic regressions, show similar results and attribute this more to differences in coefficients rather than endowments. This may point to discrimination, since the differential is not explained by endowment differences and coefficients for native Germans are consistently and significantly higher.

### 3.3 Return migration

The nature of return migration affects wage differentials over time. Short-run and long-run differentials vary based on the selection of return migrants. However, long-run returns to experience and wage time trends may also differ between native German and foreign workers. The above regressions take the entire work histories rather than specific periods of time, and the differentials are thus a weighted average of short- and long-run differentials. To examine the impact of return migration, I focus on specific periods of time. This leads to two propositions:

*Proposition 1: If return migrants are positively selected (i.e., drawn from the upper end of the wage distribution), then the relative wage position of foreigners worsens over time ceteris paribus.*

*Proposition 2: If returns to experience/wage trajectories are higher for native German workers, then the relative wage position of foreigners worsens over time ceteris paribus.*

Note that neither proposition implies the other and results could be ambiguous. Proposition 2 does follow from the Oaxaca-Blinder results: native German workers do experience a higher return to experience. Figure III.1 shows the wages at different levels of experience, where experience is defined as the number of years since the first entry into the German labor market. This figure does provide further evidence for greater returns to experience for native Germans at higher levels. In particular, native German wages increase dramatically around 15 years of experience, which could correspond to career moves into management, bonuses, and seniority-based pay.[4] Nevertheless, return migration does confound the differences between native German, foreign, and German*foreign (i.e., both) and especially at lower levels of experience.

I disentangle the impact of return migration from differences to experience by focusing on workers who first appear in the data in 1976 or later. This removes potential problems with older workers who began working prior to 1975, when the data begin. I run a simple Mincer-style log wage equation similar to Equation 1. However, here the coefficients of interest are experience and the two interaction terms of experience with foreign and German*foreign. Table III.6 shows coefficients for experience and experience squared as well as the interactions.

The experience coefficient for all workers is positive and significant, which accords with the results from Table III.5. The interaction with foreign-only is also positive and significant, but the difference between the two is small and not different than zero. The interaction with German*foreign (i.e., both) is also positive. The difference is significant at the 10% level. Experience squared is negative for all three, which indicates diminishing returns to experience. However, it is more negative for both foreign-only and both. The difference between all and foreign-only is significant at the 10% level, and the difference between all and German*foreign is significant at the 1% level. This means that at low levels of experience, the wage gains to experience are similar for native German-only and foreign-only workers. However, diminishing returns hit foreign-only workers faster, and so native German workers do gain more from experience.

Next, I compare summary statistics for any-foreign workers based on their duration in Germany. I exclude workers from the last years of data who worked under five years. I divide between those who stayed at last five years and those who left Germany–or at least the data–within this time frame. Table III.7 reports education levels for the two groups. The human capital is higher for the leavers, not the stayers; a higher proportion of leavers have reported education (over 50% vs under 36%). The picture is stark for technical college and university graduates: over 11% for leavers vs under 6% for stayers. Thus, it appears that return migrants are positively selected, i.e., they come from the upper end of the wage distribution based on the fact that they have higher human capital.

Together, these provide support for Propositions 1 and 2. Native German workers have higher

[4] I thank a staff member of the FDZ for suggesting this possibility.

returns to experience and the (potentially) highest earners from the any-foreign population are more likely to leave rather than stay. These two conclusions support an increasing wage differential between native German and foreign (and German*foreign) workers over time.

## 4 BASiD

To further examine the impact of return migration, I turn to pension data. Several differences exist between SIAB and BASiD based on available variables. As mentioned, BASiD does not contain the proportion of workers who are German. Additionally, the local establishment type is not consistent across years, so I omit it. However, several additional covariates emerge as useful. First, both West German and East German (DDR) individuals are included. I run the specification using workers with only West German/unified German work histories in order to be consistent with SIAB. Second, because BASiD samples the population differently, I include dummy variables for decade of entry in the data (2000s is the omitted variable).

I run regressions using the BASiD data using a similar form as Equation 1.

$$\log(daily\_wage_{it}) = \beta_0 + \beta_1 I_{foreign;i,t} + \beta_2 I_{German*foreign;i,t} + \beta_3 I_{1950s;i} + \beta_4 I_{1950s*foreign;i}$$

$$+ \beta_5 I_{1950s*German*foreign;i} + \cdots + \Gamma_1 Z_{i,t} + \Gamma_2 Year + \Gamma_3 Geo + \epsilon_{it} \tag{2}$$

Thus, this variable is measured only with an $i$ subscript as it does not vary over $t$. Differences across time in the interacted decade coefficients indicate changes in the differentials due to return migration.

Table III.8 shows the coefficients for decade of entry from the full specification. (The other coefficients are not reported for space.) As anticipated, the coefficients in column 1 are positive and significant; this reflects the longer working career of people who entered earlier and and the positive returns to experience. The returns to experience flatten out for foreign workers. Adding the coefficients in columns 1 and 2 for a given year reflects the effect of appearing in a decade plus being foreign. Foreign-only workers who appear in all the decades earned roughly 20% more than foreigners from the 2000s. In contrast, the returns to experience are over 40% higher for workers who appeared in the 1980s or earlier and 34% for those in the 1990s than those who first appear in the 2000s. This is consistent with Table III.6, in which diminishing returns to experience occur at lower levels of experience for foreign workers.

Column 3 provides an interesting picture of assimilation for German*foreign workers. Workers who are German*foreign who entered in earlier decades resemble native Germans. This is especially striking for those who appeared first in the 1950s and for whom the coefficients on foreign and German*foreign effectively cancel out. However, the assimilation effect weakens for later decades. German*foreign workers who appeared first in the 1960s earn 10% more than comparable foreign workers (significant at the 5% level), but they still earn less than native German workers who appear in that decade. By the 1990s, the coefficient on German*foreign is almost zero and not significant; foreign and German*foreign workers appear similar in the decadal 'bonus' that they receive.

## 5 Discussion

The two data sets provide evidence about wage differentials in West Germany/Germany. The preferred specification utilizes local labor markets and uses the 'always-random' SIAB data. Results with states rather than local labor markets (not shown) are robust. While there is some theoretical justification for using states as the geographic measure, local labor markets capture more information about the conditions and are econometrically more justifiable.

Summary statistics indicate the existence of wage differentials between German-only workers and the two comparison groups, foreign-only and German*foreign. This is consistent with the prior literature on differentials in Germany. These differences appear to be driven by differences in human capital, differential returns to human capital, and differential returns to experience. As seen in Table III.3, the coefficient on foreign is actually positive and significant. However, this masks the fact that most native German workers are in higher-education, higher-paying brackets. This accords with both Pischke (1992) and Schmidt (1997). Furthermore, the Oaxaca-Blinder regressions in Table III.5 provide evidence on how these differentials operate both through different endowments and different returns. Interestingly, foreign workers reap higher rewards from local labor market differences. This could reflect sorting into higher-paying markets, particularly urban ones.

Return migration affects wage differentials because of who returns, but differential returns potentially confound analyzing return migration. By focusing on a selected subsample, I find that returns to experience differ between native German and foreign workers, with a benefit to native German workers, but the difference matters more at higher levels. Return migration contributes to the widening of the wage differential over time because the higher-educated foreign workers are more likely to leave rather than remain.

The results from BASiD bolster these two conclusions. Workers who appear early in the data do earn more, which is consistent with positive returns to experience. However, there is a widening gap between native German and foreign workers over time; this is also consistent with return migration by the top of the distribution.

Extrapolating from my results, in the Borjas and Bratsberg framework, the sending countries are sending the lower end of their distributions. This, however, cannot be definitively shown without comparing to the distributions of workers in those sending countries.

## 6 Conclusion

The analysis of these two large administrative data sets from Germany provides answers to several key questions. First, per SIAB, there do appear to be wage differentials between Germans and foreigners. These differentials diminish with the inclusion of additional control variables but do not disappear. Using local-labor-market characteristics provides more precise estimates and deals with potential omitted variable bias.

Second, these differentials appear to be the result of both human capital differences and differential returns to experience. Local labor markets actually mitigate the differences due to sorting by foreign workers.

Third, return migration exacerbates this differential. The highest-educated foreign workers leave, which reduces both the level and presumably the trajectory of foreign wages in the aggregate.

There are several caveats to this. First, human capital may be mismeasured or misspecified; this is especially a problem for foreign workers who leave early. Interestingly, foreign workers who leave early are more likely to have a degree reported, which reduces this impact, but roughly half of them do not have information. Second, mortality may be a factor, but that posits that differential mortality affects different parts of the distribution for German-only workers and of either foreign group. This seems implausible given the positive correlation between health and other aspects of human capital. However, it may occur and thus may affect any differential between foreign and German workers.

The difference may also result from differential participation in the labor force in addition to migration. Those immigrants who end up working may differ from those who choose not to work. While this may be inappropriate in a context without social benefits (and even deportation) for non-working immigrants, such as the United States, this may not be appropriate for West Germany/Germany. Further work on this using benefit receipt helps to deal with the first-stage selection into work.

Nationality is a vague term that has potential problems. While in theory it is an objective measure due to citizenship status, in contrast with self-reported ethnicity/race, in reality it is also hard to distinguish between different groups cleanly. I attempt to overcome this using foreign-only and German*foreign groups, but in the latter, I cannot distinguish between naturalized citizens and German-born workers whose status depends on jus sanguinis. As pointed out in the discussion, switching over time may be important. Further research based on switching and using pooled groups may assist.

Finally, experience in the home country prior to immigration is impossible to measure. To bypass this, I use age in my main specifications. However, age is an imperfect proxy. Further research using the GSOEP or other qualitative data may address how large a problem this is.

Overall, the method of comparing similar data sets and focusing on the difference in sampling bypasses problems of not directly observing return migration. While the above caveats serve as a caution for an overzealous interpretation, the results here point to ambiguous selection effects on 'remainers' in the German labor force that go against theory.

## References

[1] Bartolucci, Cristian (2009), "Understanding the Native-Immigrant Wage Gap using Matched Employer-Employee data. Evidence from Germany," working paper.

[2] Becker, Gary S. (1957), *The economics of discrimination*, Chicago: University of Chicago Press.

[3] Borjas, George J., and Bernt Bratsberg (1996), "Who Leaves? The Outmigration of the Foreign-Born," *The Review of Economics and Statistics* 78 (1): 165-176.

[4] Borjas, George J. (1987), "Self-Selection and the Earnings of Immigrants," *The American Economic Review* 77 (4): 531-553.

[5] Börsch-Supan, Axel and Christina Wilke (2004), "The German public pension system: How it was, how it will be," NBER working paper 10525.

[6] Chin, Rita (2007), *The guest worker question in postwar Germany*, Cambridge: Cambridge University Press.

[7] D'Amuri, Francesco, Gianmarco Ottaviano, and Giovanni Peri (2010), "The labor market impact of immigration in Western Germany in the 1990s," *European Economic Review*, 54 (4): 550-570.

[8] Dustmann, Christian (1997a), "Differences in the labor market behavior between temporary and permanent migrant women," *Labour Economics*, 4: 29-46.

[9] Dustmann, Christian (1997b), "Return migration, uncertainty and precautionary savings," *Journal of Development Economics* 52 (2): 295-316.

[10] Federal Statistical Office and the statistical Offices of the Länder (2014), "Area and population - Foreign population," accessed 16 April 2015 at http://www.statistik-portal.de/statistik-portal/en/en_jb01_jahrtab2.asp.

[11] Finnish Centre for Pensions (2009), "Pension contribution level in Germany," official report.

[12] Fitzenberger, Bernd, Aderonke Osikominu, and Robert Völter, (2006), "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 126 (3): 405-436.

[13] Galor O, and O Stark (1991) "The probability of return migration, migrants' work effort, and migrants' performance," *Journal of Development Economics* 35 (2): 399-405.

[14] Hochfellner, Daniela, Dana Müller, and Anja Wurdack (2012), "Biographical data of social insurance agencies in Germany. Improving the content of administrative data," in *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*, Bd. 132, H. 3, S. 443-451.

[15] Hochfellner, Daniela, Dana Müller, and Anja Wurdack (2011) "BASiD - Biografiedaten ausgewhlter Sozialversicherungstrger in Deutschland," FDZ-Datenreport, 09/2011, Nürnberg.

[16] Kropp, Per and Schwengler, Barbara (2012), "Vergleich von funktionalen Arbeitsmarktabgrenzungen in Deutschland," *Sozialer Fortschritt* 61 (1): 11-20.

[17] Merkle L, and KF Zimmermann (1992), "Savings, remittances, and return migration," *Economics Letters* 38 (1): 77-81.

[18] Pischke, Jörn-Steffen (1992), "Assimilation and the earnings of guestworkers in Germany," ZEW Discussion Papers, No. 92-17.

[19] Pischke, Jörn-Steffen and Johannes Velling (1997), "Employment effects of immigration to Germany: An anlysis based on local labor markets," *The Review of Economics and Statistics*, 79 (4): 594-604.

[20] Schmidt, Christoph (1997), "Immigrant performance in Germany: Labor earnings of ethnic German migrants and foreign guest-workers," *The Quarterly Review of Economics and Finance*, 37: 379-397.

[21] Vasileva, Katya (2011), "Population and social conditions," Eurostat: Statistics in Focus 34/2011, accessed 1 May 2014 at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-11-034/EN/KS-SF-11-034-EN.PDF.

[22] Velling, Johannes (1995), "Wage discrimination and occupational segregation of foreign male workers in Germany," ZEW Discussion Papers, No. 95-04.

[23] vom Berge, Philipp, Marion König, and Stefan Seth (2013), "Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2010," FDZ Datenreport, 01/2013 (de), Nürnberg.

[24] Yang, Dean (2006), Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants Responses to Exchange Rate Shocks, *Review of Economics and Statistics* 88(4), pp. 715-735.

Figure III.1: Wage trajectories

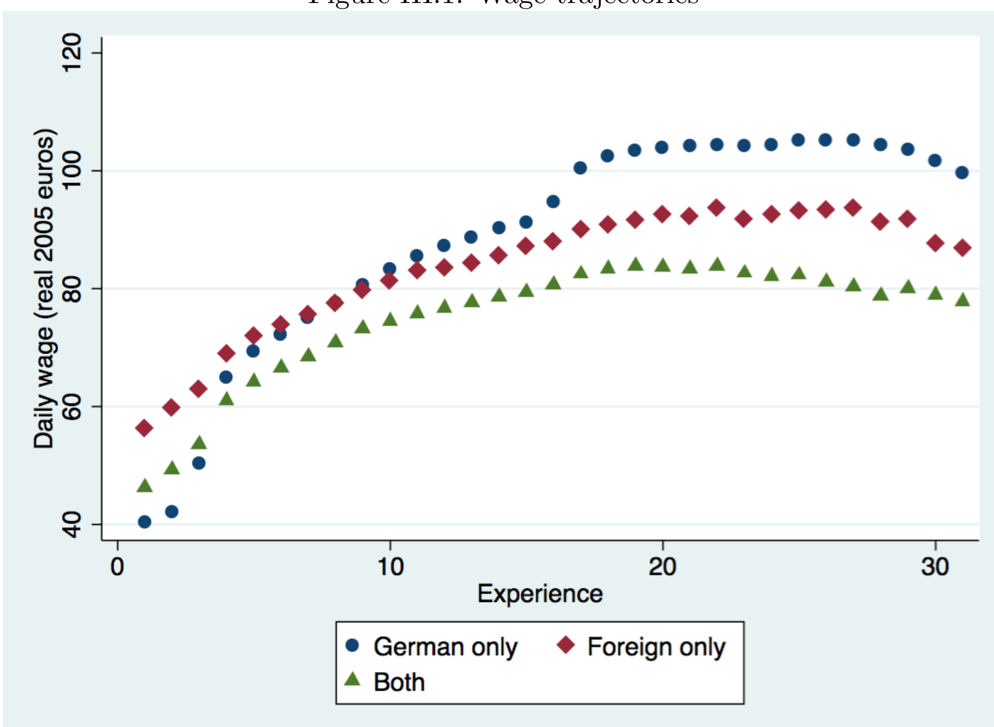


Table III.1: Descriptive statistics (individual level)

| | SIAB | | | | BASiD | | | |
|---|---|---|---|---|---|---|---|---|
| | All | German | Foreign | Both | All | German | Foreign | Both |
| Only German | 0.88 | | | | 0.86 | | | |
| | (0.0003) | | | | (.0002) | | | |
| Only foreign | 0.06 | | | | 0.04 | | | |
| | (0.0002) | | | | (.0009) | | | |
| German*foreign | 0.06 | | | | 0.10 | | | |
| | (0.0002) | | | | (.0001) | | | |
| Female | 0.46 | 0.46 | 0.35 | 0.43 | 0.45 | 0.45 | 0.34 | 0.44 |
| | (0.0004) | (0.0005) | (0.002) | (0.002) | (0.0002) | (0.003) | (0.008) | (0.005) |
| Max education: | | | | | | | | |
| No degree | 0.19 | 0.15 | 0.51 | 0.31 | 0.10 | 0.07 | 0.45 | 0.20 |
| | (0.0003) | (0.0003) | (0.002) | (0.002) | (0.0001) | (0.0001) | (0.0008) | (0.0004) |
| Only voc training | 0.61 | 0.64 | 0.37 | 0.51 | 0.71 | 0.73 | 0.45 | 0.58 |
| | (0.0004) | (0.0005) | (0.002) | (0.002) | (0.0004) | (0.0002) | (0.0002) | (0.0007) |
| HS degree, no training | 0.02 | 0.02 | 0.02 | 0.03 | 0.01 | 0.01 | 0.01 | 0.01 |
| | (0.0004) | (0.0001) | (0.001) | (0.006) | (0.00004) | (0.00004) | (0.0002) | (0.001) |
| HS degree, training | 0.07 | 0.07 | 0.03 | 0.07 | 0.08 | 0.08 | 0.03 | 0.08 |
| | (0.0002) | (0.0002) | (0.006) | (0.009) | (0.0001) | (0.0001) | (0.0003) | (0.003) |
| University | 0.07 | 0.07 | 0.06 | 0.06 | 0.07 | 0.07 | 0.04 | 0.08 |
| | (0.0002) | (0.0002) | (0.0008) | (0.0008) | (0.0001) | (0.0001) | (0.0003) | (0.0003) |
| Observations | 1278072 | 1120826 | 75720 | 81526 | 381743 | 244835 | 58178 | 78730 |

Source: See text. Standard errors shown in parentheses. BASiD calculations are weighted per the IAB. Column note: 'All' is the entire sample. 'German' is for individuals reported only as German, 'Foreign' for individuals reported only as foreign, and 'Both' for individuals reported as both over their histories.

Table III.2: Descriptive statistics (spell level)

| | SIAB | | | | BASiD | | | |
|---|---|---|---|---|---|---|---|---|
| | All | German | Foreign | Both | All | German | Foreign | Both |
| Real daily wage | | | | | | | | |
| All | 70.55 | 71.07 | 69.91 | 62.73 | 70.37 | 70.88 | 69.84 | 66.36 |
| | (0.01) | (0.01) | (0.04) | (0.04) | (0.02) | (0.02) | (0.05) | (0.04) |
| Males | 83.07 | 84.06 | 77.56 | 72.36 | 82.69 | 83.765 | 78.87 | 76.51 |
| | (0.01) | (0.01) | (0.05) | (0.05) | (0.02) | (0.03) | (0.07) | (0.06) |
| Age at spell | 38.39 | 38.54 | 38.51 | 35.88 | 35.56 | 35.68 | 35.45 | 34.65 |
| | (0.003) | (0.003) | (0.01) | (0.01) | (.005) | (.006) | (0.02) | (0.01) |
| % German firm | 92.32 | 94.12 | 70.73 | 77.50 | | | | |
| | (.3) | (.2) | (3.0) | (2.4) | | | | |
| N | 16156447 | 14652139 | 598325 | 905983 | 4481051 | 3217299 | 405972 | 857780 |
| *Male obs* | 9257598 | 8286170 | 415067 | 556361 | 2394722 | 1698233 | 240203 | 456286 |

Source: See text. Standard errors shown in parentheses. BASiD calculations are weighted per the IAB. Column note: 'All' is the entire sample. 'German' is for individuals reported only as German, 'Foreign' for individuals reported only as foreign, and 'Both' for individuals reported as both over their histories.

Table III.3: Native vs foreign

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Foreign | 0.241*** | 0.202*** | 0.0456*** | 0.171*** | 0.197*** | 0.0443*** | 0.0556*** |
| | [0.00283] | [0.00266] | [0.00181] | [0.00245] | [0.00249] | [0.00170] | [0.00172] |
| Foreign*German | -0.210*** | -0.150*** | -0.0649*** | -0.115*** | -0.119*** | -0.0512*** | -0.0528*** |
| | [0.00416] | [0.00382] | [0.00247] | [0.00350] | [0.00347] | [0.00232] | [0.00230] |
| Age | 0.116*** | 0.114*** | 0.0617*** | 0.109*** | 0.109*** | 0.0596*** | 0.0596*** |
| | [0.000259] | [0.000239] | [0.000178] | [0.000237] | [0.000237] | [0.000176] | [0.000176] |
| Age sq | -0.00136*** | -0.00133*** | -0.000678*** | -0.00127*** | -0.00127*** | -0.000655*** | -0.000656*** |
| | [0.00000338] | [0.00000313] | [0.00000226] | [0.00000310] | [0.00000310] | [0.00000225] | [0.00000225] |
| Female | -0.467*** | -0.437*** | -0.287*** | -0.380*** | -0.382*** | -0.270*** | -0.270*** |
| | [0.00101] | [0.000960] | [0.000804] | [0.00104] | [0.00104] | [0.000800] | [0.000800] |
| Voc training degree | 0.370*** | 0.401*** | 0.0677*** | 0.417*** | 0.415*** | 0.0724*** | 0.0718*** |
| | [0.00144] | [0.00135] | [0.00105] | [0.00131] | [0.00131] | [0.00103] | [0.00103] |
| HS degree, no training | -0.133*** | -0.0497*** | -0.00569** | -0.0421*** | -0.0432*** | -0.0133*** | -0.0132*** |
| | [0.00536] | [0.00486] | [0.00285] | [0.00479] | [0.00479] | [0.00279] | [0.00279] |
| HS degree, voc training | 0.450*** | 0.534*** | 0.108*** | 0.549*** | 0.546*** | 0.108*** | 0.108*** |
| | [0.00264] | [0.00253] | [0.00179] | [0.00247] | [0.00247] | [0.00175] | [0.00175] |
| Technical college | 0.681*** | 0.740*** | 0.206*** | 0.741*** | 0.739*** | 0.205*** | 0.204*** |
| | [0.00257] | [0.00245] | [0.00194] | [0.00242] | [0.00242] | [0.00191] | [0.00191] |
| University | 0.724*** | 0.801*** | 0.272*** | 0.824*** | 0.822*** | 0.276*** | 0.276*** |
| | [0.00240] | [0.00236] | [0.00194] | [0.00234] | [0.00234] | [0.00192] | [0.00192] |
| Voc training degree, foreign | -0.276*** | -0.290*** | -0.0679*** | -0.276*** | -0.276*** | -0.0659*** | -0.0655*** |
| | [0.00424] | [0.00406] | [0.00281] | [0.00379] | [0.00373] | [0.00262] | [0.00259] |
| Voc training degree, foreign*German | 0.0661*** | 0.0816*** | 0.0122*** | 0.0699*** | 0.0651*** | 0.00670* | 0.00438 |
| | [0.00596] | [0.00558] | [0.00369] | [0.00522] | [0.00515] | [0.00346] | [0.00342] |
| HS degree, no training, foreign | -0.224*** | -0.181*** | -0.0708*** | -0.142*** | -0.146*** | -0.0605*** | -0.0620*** |
| | [0.0284] | [0.0246] | [0.0149] | [0.0234] | [0.0232] | [0.0142] | [0.0141] |
| HS degree, no training, foreign*German | 0.0530 | 0.0182 | 0.000342 | -0.00266 | -0.00618 | -0.00507 | -0.00683 |
| | [0.0334] | [0.0291] | [0.0177] | [0.0279] | [0.0277] | [0.0170] | [0.0169] |
| HS degree, voc training, foreign | -0.263*** | -0.276*** | -0.0618*** | -0.237*** | -0.242*** | -0.0534*** | -0.0555*** |
| | [0.0181] | [0.0170] | [0.0112] | [0.0164] | [0.0162] | [0.0109] | [0.0108] |
| HS degree, voc training, foreign*German | 0.0194 | 0.0201 | -0.0206 | -0.00845 | -0.0123 | -0.0304** | -0.0323*** |
| | [0.0212] | [0.0198] | [0.0129] | [0.0191] | [0.0189] | [0.0126] | [0.0125] |
| Technical college, foreign | -0.229*** | -0.246*** | -0.0602*** | -0.228*** | -0.238*** | -0.0685*** | -0.0727*** |
| | [0.0174] | [0.0167] | [0.0113] | [0.0165] | [0.0163] | [0.0113] | [0.0112] |
| Technical college, foreign*German | 0.0428* | 0.0404* | 0.000931 | 0.0296 | 0.0268 | 0.00359 | 0.00206 |
| | [0.0234] | [0.0222] | [0.0143] | [0.0219] | [0.0217] | [0.0142] | [0.0141] |
| University, foreign | -0.230*** | -0.242*** | -0.0718*** | -0.212*** | -0.226*** | -0.0704*** | -0.0759*** |
| | [0.0112] | [0.0110] | [0.00685] | [0.0108] | [0.0107] | [0.00680] | [0.00677] |
| University, foreign*German | 0.0955*** | 0.0887*** | 0.0216** | 0.0612*** | 0.0613*** | 0.0138 | 0.0137 |
| | [0.0151] | [0.0145] | [0.00916] | [0.0141] | [0.0140] | [0.00905] | [0.00900] |
| Proportion native workers | | | | | 0.125*** | | 0.0578*** |
| | | | | | [0.00350] | | [0.00252] |
| Geo, year | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm cat | No | No | No | Yes | Yes | Yes | Yes |
| Worker cat | No | No | Yes | No | No | Yes | Yes |
| Observations | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 |

Standard errors in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table III.4: Native vs foreign (disaggregated)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Turkey | 0.222*** | 0.192*** | 0.0528*** | 0.147*** | 0.171*** | 0.0444*** | 0.0547*** |
| | [0.00406] | [0.00383] | [0.00247] | [0.00353] | [0.00353] | [0.00232] | [0.00233] |
| Turkey*Germany | -0.138*** | -0.0867*** | -0.0277*** | -0.0565*** | -0.0580*** | -0.0156*** | -0.0161*** |
| | [0.00655] | [0.00606] | [0.00370] | [0.00558] | [0.00553] | [0.00350] | [0.00347] |
| Italy | 0.281*** | 0.228*** | 0.0474*** | 0.204*** | 0.229*** | 0.0502*** | 0.0610*** |
| | [0.00651] | [0.00613] | [0.00442] | [0.00564] | [0.00562] | [0.00416] | [0.00412] |
| Italy*Germany | -0.142*** | -0.117*** | -0.0542*** | -0.0887*** | -0.0912*** | -0.0422*** | -0.0432*** |
| | [0.00988] | [0.00928] | [0.00655] | [0.00856] | [0.00848] | [0.00617] | [0.00610] |
| Greece | 0.310*** | 0.246*** | 0.0667*** | 0.182*** | 0.207*** | 0.0498*** | 0.0605*** |
| | [0.00814] | [0.00760] | [0.00542] | [0.00657] | [0.00646] | [0.00488] | [0.00480] |
| Greece*Germany | -0.147*** | -0.114*** | -0.0654*** | -0.0813*** | -0.0816*** | -0.0513*** | -0.0514*** |
| | [0.0144] | [0.0135] | [0.00929] | [0.0118] | [0.0116] | [0.00840] | [0.00828] |
| Poland | -0.106*** | -0.0172 | -0.116*** | 0.105*** | 0.142*** | -0.0702*** | -0.0535*** |
| | [0.0260] | [0.0237] | [0.0191] | [0.0226] | [0.0225] | [0.0177] | [0.0174] |
| Poland*Germany | -0.0951*** | -0.106*** | 0.0289 | -0.197*** | -0.220*** | 0.00363 | -0.00720 |
| | [0.0334] | [0.0304] | [0.0222] | [0.0289] | [0.0288] | [0.0208] | [0.0205] |
| Yugoslavia | 0.268*** | 0.211*** | 0.0585*** | 0.192*** | 0.213*** | 0.0624*** | 0.0718*** |
| | [0.00617] | [0.00583] | [0.00396] | [0.00520] | [0.00518] | [0.00363] | [0.00361] |
| Yugoslavia*Germany | -0.184*** | -0.131*** | -0.0421*** | -0.108*** | -0.112*** | -0.0355*** | -0.0373*** |
| | [0.0103] | [0.00948] | [0.00603] | [0.00862] | [0.00856] | [0.00560] | [0.00556] |
| Other foreign | 0.125*** | 0.115*** | 0.000941 | 0.116*** | 0.141*** | 0.0114*** | 0.0222*** |
| | [0.00633] | [0.00576] | [0.00368] | [0.00526] | [0.00525] | [0.00345] | [0.00344] |
| Other foreign*Germany | -0.251*** | -0.186*** | -0.0913*** | -0.156*** | -0.163*** | -0.0808*** | -0.0837*** |
| | [0.00893] | [0.00805] | [0.00518] | [0.00744] | [0.00739] | [0.00489] | [0.00485] |
| Age | 0.116*** | 0.115*** | 0.0618*** | 0.109*** | 0.110*** | 0.0597*** | 0.0597*** |
| | [0.000259] | [0.000239] | [0.000178] | [0.000237] | [0.000237] | [0.000176] | [0.000176] |
| Age sq | -0.00137*** | -0.00133*** | -0.000679*** | -0.00127*** | -0.00127*** | -0.000656*** | -0.000657*** |
| | [0.00000337] | [0.00000313] | [0.00000226] | [0.00000310] | [0.00000310] | [0.00000225] | [0.00000225] |
| Female | -0.467*** | -0.437*** | -0.287*** | -0.381*** | -0.382*** | -0.270*** | -0.270*** |
| | [0.00101] | [0.000960] | [0.000805] | [0.00104] | [0.00104] | [0.000801] | [0.000800] |
| Voc training degree | 0.368*** | 0.400*** | 0.0670*** | 0.416*** | 0.414*** | 0.0717*** | 0.0710*** |
| | [0.00143] | [0.00134] | [0.00105] | [0.00130] | [0.00130] | [0.00102] | [0.00102] |
| HS degree, no training | -0.135*** | -0.0513*** | -0.00635** | -0.0435*** | -0.0447*** | -0.0139*** | -0.0138*** |
| | [0.00535] | [0.00485] | [0.00285] | [0.00478] | [0.00478] | [0.00279] | [0.00279] |
| HS degree, voc training | 0.448*** | 0.532*** | 0.107*** | 0.547*** | 0.544*** | 0.107*** | 0.107*** |
| | [0.00264] | [0.00252] | [0.00179] | [0.00246] | [0.00246] | [0.00175] | [0.00175] |
| Technical college | 0.679*** | 0.738*** | 0.205*** | 0.739*** | 0.737*** | 0.204*** | 0.203*** |
| | [0.00257] | [0.00245] | [0.00194] | [0.00242] | [0.00242] | [0.00191] | [0.00191] |
| University | 0.721*** | 0.799*** | 0.271*** | 0.822*** | 0.820*** | 0.275*** | 0.275*** |
| | [0.00239] | [0.00235] | [0.00194] | [0.00233] | [0.00234] | [0.00191] | [0.00191] |
| Proportion native workers | | | | | 0.122*** | | 0.0566*** |
| | | | | | [0.00350] | | [0.00252] |
| Geo, year | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm cat | No | No | No | Yes | Yes | Yes | Yes |
| Worker cat | No | No | Yes | No | No | Yes | Yes |
| Observations | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 | 16156447 |

Standard errors in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table III.5: German vs Any foreign

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| overall | | | | | | |
| Native | 4.038*** | 4.038*** | 4.038*** | 4.038*** | 4.038*** | 4.038*** |
| | [0.000422] | [0.000526] | [0.000441] | [0.000441] | [0.000533] | [0.000534] |
| Any foreign | 3.970*** | 3.970*** | 3.970*** | 3.970*** | 3.970*** | 3.970*** |
| | [0.00119] | [0.00149] | [0.00134] | [0.00136] | [0.00154] | [0.00156] |
| difference | 0.0681*** | 0.0681*** | 0.0681*** | 0.0681*** | 0.0681*** | 0.0681*** |
| | [0.00126] | [0.00158] | [0.00141] | [0.00143] | [0.00163] | [0.00165] |
| endowments | -0.00244* | 0.0212*** | -0.0214*** | 0.0480*** | 0.0296*** | 0.0759*** |
| | [0.00126] | [0.00158] | [0.00142] | [0.00143] | [0.00163] | [0.00164] |
| coefficients | 0.0169*** | 0.0250*** | 0.0213*** | 0.0518*** | 0.0198*** | 0.0506*** |
| | [0.000425] | [0.000197] | [0.000503] | [0.000599] | [0.000230] | [0.000368] |
| interaction | 0.0536*** | 0.0219*** | 0.0682*** | -0.0317*** | 0.0187*** | -0.0584*** |
| | [0.000442] | [0.000208] | [0.000526] | [0.000618] | [0.000244] | [0.000377] |
| Geo, year | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm cat | No | No | Yes | Yes | Yes | Yes |
| Worker cat | No | Yes | No | No | Yes | Yes |
| Observations | 16102275 | 16102275 | 16102275 | 16102275 | 16102275 | 16102275 |

Standard errors in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table III.6: Returns to experience

| | All | Foreign-only | Foreign*native |
|---|---|---|---|
| Experience | 0.115*** | 0.004 | 0.0032* |
| | (0.002) | (0.002) | (0.002) |
| Experience$^2$ | -0.000723*** | -0.000436* | -0.0011*** |
| | (0) | (0) | (0) |
| N (total) | | 8872834 | |

Clustered standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table III.7: Comparative statistics (spell level)

| | Left | Stayed |
|---|---|---|
| No degree | 49.48 | 63.77 |
| Voc training | 32.83 | 24.90 |
| HS degree, no training | 3.64 | 3.64 |
| HS degree+training | 2.72 | 2.14 |
| Tech college | 1.95 | 1.22 |
| University | 9.37 | 4.33 |
| N | 12021 | 72528 |

Table III.8: Decade of entry

| | (1) | (2) | (3) |
|---|---|---|---|
| | All | Foreign | Foreign*native |
| Appeared first in 1950s | 0.453*** | -0.266*** | 0.224*** |
| | [0.0144] | [0.0521] | [0.0629] |
| Appeared first in 1960s | 0.426*** | -0.194*** | 0.104** |
| | [0.0125] | [0.0373] | [0.0446] |
| Appeared first in 1970s | 0.411*** | -0.179*** | 0.0742* |
| | [0.0106] | [0.0334] | [0.0395] |
| Appeared first in 1980s | 0.438*** | -0.209*** | 0.0744** |
| | [0.00878] | [0.0290] | [0.0348] |
| Appeared first in 1990s | 0.343*** | -0.128*** | 0.00945 |
| | [0.00652] | [0.0246] | [0.0299] |
| Observations | | 4470943 | |
| FE | | Year, Geo, Worker type | |

Standard errors in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$